

REVIEW

Open Access



Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: a scoping review

Christian W. Mayer^{1*} , Andreas Rausch¹ and Jürgen Seifried¹

*Correspondence:
christian.mayer@uni-mannheim.de

¹Business School, Economic
and Business Education,
University of Mannheim, L4, 1,
68161 Mannheim, Germany

Abstract

Recently, many studies have been published on the use of eye-tracking to analyse complex problem-solving processes within authentic computer-based learning and training environments. This scoping review aims to provide a systematic report of the current state-of-the-art for related papers. Specifically, this work offers a scoping review of studies that analyse problem-solving processes by using eye-tracking (alongside additional process data such as log files, think aloud, facial expression recognition algorithms, or psychophysiological measures) within authentic technology-based learning and training environments for professional and vocational education and training (VET). A total of 12 studies were identified. The most commonly calculated measures in eye-tracking research are position measures, and these are almost exclusively position duration measures such as the proportion of fixation times or total dwell times. Count measures are also mostly related to the number or proportion of fixations and dwells. Movement measures are rarely computed and usually refer to saccade directions or a scan path. Also, latency and distance measures are almost never calculated. Eye-tracking data is most often analysed for group comparisons between experts vs. novices or high vs. low-performing groups by using common statistical methods such as t-test, (M)ANOVA, or non-parametric Mann–Whitney-U. Visual attention patterns in problem-solving are examined with heat map analyses, lag sequential analyses, and clustering. Recently, linear mixed-effects models have been applied to account for between and within-subjects differences. Also, post-hoc performance predictions are being developed for future integration into multimodal learning analytics. In most cases, self-reporting is used as an additional measurement for data triangulation. In addition to eye-tracking, log files and facial expression recognition algorithms are also used. Few studies use shimmer devices to detect electrodermal activity or practice concurrent thinking aloud. Overall, Haider and Frensch's (1996, 1999) "information reduction hypothesis" is supported by many studies in the sample. High performers showed a higher visual accuracy, and visual attention was more focused on relevant areas, as seen by fewer fixation counts and higher fixation duration. Low performers showed significantly fewer fixation durations or substantially longer fixation durations and less selective visual attention. Performance is related to prior knowledge and differences in

cognitive load. Eye-tracking, (in combination with other data sources) may be a valid method for further research on problem-solving processes in computer-based simulations, may help identify different patterns of problem-solving processes between performance groups, and may hold additional potential for individual learning support.

Keywords: Complex problem-solving, Computer-based learning environments, Computer-based simulations, Vocational education and training, VET, Online measurements, Eye-tracking

Introduction

In educational research, collecting behavioural data is becoming increasingly important to learn more about cognitive and metacognitive processes during learning and instruction. Eye-tracking, a method for analysing the course of gaze, is increasingly used in educational research to improve the instructional design in computer-based learning environments and multimedia learning, to understand and promote the development of expertise, and to visualize the eye movements of experts (Jarodzka et al. 2017). Systematic reviews and meta-analyses of eye-tracking studies are available for various domains (e.g., medical education: Ashraf et al. 2018; mathematics: Strohmaier et al. 2020). Similarly, reviews on multimedia learning and instructional design were conducted (e.g., Alemdag and Cagiltay 2018: multimedia learning; Yang et al. 2018: instructional design of e-learning). However, little research has been done on vocational education and training (VET). This is especially true when more complex vocational tasks are the focus. Therefore, the paper at hand provides an overview of studies that have analysed domain-specific problem-solving processes by using eye-tracking (next to additional online data such as log files or psychophysiological measures) within authentic computer-based learning and training environments in professional training or vocational education and training. The review of the current state of research is conducted as a scoping review. Scoping reviews (Arksey and O'Malley 2005) are considered a useful approach to examining the design and conduct of research on a particular topic (for the key features of the scoping review approach see Munn et al. 2018). Similar to systematic reviews, scoping reviews are transparent and replicable by following a rigorous study search and selection process. Because fewer studies were identified that followed the same study design, we focused on a scoping review.

While systematic reviews aim to answer a specific question, a scoping review identifies, reports, and discusses a broader perspective on a given topic (Arksey and O'Malley 2005; Munn et al. 2018; Van Ostaeyen et al. 2022), such as analysing domain-specific problem-solving in computer-based simulations by using eye-tracking.

Computer-based learning environments (CBLE) refer to a broad range of technologies to support learning and instruction (Lajoie and Naismith 2012). This review includes studies that used eye-tracking as the primary method for collecting behavioural data in computer simulations and serious games. To control for different levels of immersion and interactivity, we excluded close-to-reality simulations, such as in situ studies and realistic simulator training (e.g., flight simulation including a full cockpit, virtual reality welding simulation, nursing practice simulation with mannequins or actors) and similarly, we excluded studies that examined cross-sectoral problem-solving testing (e.g., Raven Matrices, Tower of Hanoi). Also, we excluded studies on collaborative

problem-solving, as these studies often focus on the phenomena of joint visual attention. This rigorous procedure resulted in a sample of studies, all of which analysed domain-specific problem-solving in computer-based simulations (see Fig. 1).

The scoping review at hand aims to provide a systematic report of the current state-of-the-art in an emerging research field. It contains a variety of eye-tracking and process data measures in a broad range of domain-specific problem-solving tasks. Eye movements can be analysed with various measurements and should be carefully collected and interpreted (Holmqvist et al. 2011). In multimedia learning research, there is a wide range of empirical research methods. Online process tracking techniques such as eye tracking can be combined with other common measurement methods to draw better inferences (Jarodzka 2021). Thus, we address the following questions:

RQ1: Which eye-tracking measures and additional behavioural measurements were used and how were they analysed?

RQ2: What are the main findings of online data measures in relation to solving complex problems in computer-based simulations in VET?

The paper is structured in five sections. First, a brief overview of the theoretical background on problem-solving, computer-based simulations, and eye-tracking is given. Second, the methodological approach of this review (identification, screening, eligibility, and inclusion) is presented. Third, the main findings are reported. Finally, the results are discussed, and the limitations of this work and implications for future research are considered.

Theoretical background

Problem-solving

Problems arise when someone has a goal but lacks the knowledge of how to achieve it (Duncker 1945). Various problem types exist. Jonassen (2000) provides a taxonomy of 11 problems from well-structured such as algorithmic problems, to ill-structured such as dilemmas, in which the more ill-structured problem types may encompass more structured problems. Following Dörner (1987; Funke 2012), complex problems include various interconnected variables, multiple and conflicting goals, a lack of transparency, and dynamic development. It is difficult to accurately capture the scope of requirements to solve a particular problem. In educational contexts, problem difficulty is often

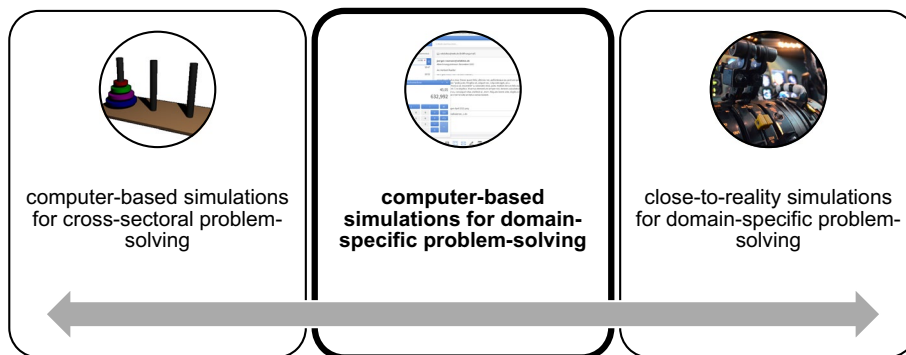


Fig. 1 The scope of this review (in bold) is illustrated within a continuum between cross-sectoral and domain-specific problem-solving simulations (pictures CC0-licensed)

assessed ex post facto based on solution rates (Jonassen and Hung 2008). Furthermore, it is challenging to define the difficulty or complexity of a problem a priori because the subjective perception of a problem varies with prior knowledge and experience in the respective domain (Dörner 1997). Similarly, Mayer and Wittrock (2006) distinguish routine and non-routine aspects of a problem, with a routine problem defined as a problem “for which the problem solver already possesses a ready-made solution procedure” (p. 288). In the context of competence assessment, Williamson et al. (2006) attempt to objectively define a task as complex if (a) the problem solver has to undergo multiple, non-trivial, domain-relevant steps and/or cognitive processes, (b) multiple features of task performance are captured, (c) task performance is relatively unconstrained, and (d) evaluations of task solutions recognise the interdependence of task features and aspects of performance.

Solving problems requires cognitive, metacognitive, and non-cognitive processes (Frensch and Funke 1995; Mayer 1998; Jonassen 2000). Weinert (2001) defines such competencies as a combination of “[...] intellectual abilities, content-specific knowledge, cognitive skills, domain-specific strategies, routines and subroutines, motivational tendencies, volitional control systems, personal value orientations, and social behaviours” (p. 51). Similarly, Fischer and Neubert (2015) define problem-solving competence as a multidimensional construct that includes knowledge, skills, abilities, and other components (KSAO), with ‘other components’ referring to non-cognitive facets such as frustration tolerance and a positive attitude in particular. Following Mayer and Wittrock (2006), problem-solving is preferably related to a specific domain instead of general heuristics (domain-specific principle), most likely restricted to a certain problem and not widely transferable to other problems (near-transfer principle), and should be integrated into teaching as guided problem-solving tasks to foster learning (knowledge integrating principle). Therefore, in order to promote problem-solving skills, problem-oriented tasks should be embedded as authentic, domain-specific scenarios in VET.

Problem-solving in computer-based simulations

Authentic domain-specific problems are so-called ‘metaproblems’ (Jonassen 2000), a combination of many problem variations and types that are connected over a single domain. Metaproblems can be illustrated within computer-based simulations to replicate real-world tasks in a safe environment for training and learning purposes while providing an authentic and dynamic simulation-based learning scenario that changes either with decisions (interactions), with time, or both (Dörner and Funke 2017). Such open-ended environments emphasize learner-centred activities, setting authentic tasks for learners, and providing them with authentic tools (Hannafin 1995; Clarebout et al. 2009). The experimental learning opportunity addresses the cognitive, motivational, affective, psychomotor, and social aspects of learning (Breckwoldt et al. 2014). Early research on problem-solving was conducted within computer-simulated microworlds (Brehmer and Dörner 1993). Nowadays, especially in the field of vocational education and training, there is a large number of domain-specific, authentic computer-based simulations to promote competence development in general and domain-specific problem-solving competence in particular (Beck et al. 2016; Rausch et al. 2016). However, most research in VET focuses on outcomes (competence assessment, learning performance,

etc.) and not on the processes that precede these outcomes (Abele 2018). Therefore, it seems worthwhile to highlight the methodological advantages of process data channels such as eye-tracking.

Eye-tracking, eye movements, and eye-tracking in problem-solving

Eye-tracking is a technology used as a research method for recording eye behaviour such as pupil dilation, blinking, and especially eye movements, as an indicator of visual attention when processing information (Holmqvist et al. 2011; Duchowski 2017; Holmqvist and Andersson 2017). Eye-tracking in computer-based simulations might help to make inferences about cognitive and metacognitive processes during learning (van Gog et al. 2009; van Gog and Jarodzka 2013).

Eye movements reflect top-down (goal-driven or endogenous) and bottom-up (stimulus-driven or exogenous) visual attention (Rayner 1998; Theeuwes 2010; Orquin and Mueller Loose 2013). Bottom-up control depends on stimulus features, such as visual saliency, i.e., the subjective quality of a stimulus that grabs visual attention (contrast, colour, movements). Top-down control depends on observer features, such as expertise, prior knowledge, tasks, etc. Also, individual eye features need to be considered. The most common types of eye movement events are fixations and saccades, where a fixation refers to the state when the eyes remain still (e.g., a stop during reading) while a saccade refers to the motion of the eyes between fixations (Holmqvist and Andersson 2017). According to the 'eye-mind hypothesis', fixations should be a proxy of cognitive processing: "the eye remains fixated on a word as long as the word is being processed" (Just and Carpenter 1980). This influential assumption (originally related to reading research, but also tested beyond reading) has been challenged several times (Underwood and Everatt 1992; Anderson et al. 2004), and today there is a consensus that visual attention somewhat precedes gaze and that overt and covert attention can differ (Holmqvist and Andersson 2017).

Eye-movement data can reveal differences in visual attention to areas of interest (AOI) during problem-solving processes. Following Haider and Frensch's 'information reduction hypothesis' (1996, 1999), deliberate practice helps students to learn to ignore redundant information and focus more on relevant information. Thus, especially experts have learned to distinguish relevant from irrelevant task-information through practice (Lee and Anderson 2001) and make use of efficient cognitive strategies through experience (van Merriënboer 2013). A meta-analysis examining the effects of expertise on visual comprehension conducted by Gegenfurtner and colleagues (2011) supported that experts have shorter fixation durations, more fixations on relevant areas, and fewer fixations on irrelevant areas than novices. Experts also showed selective attention through parafoveal processing (unattended locations of the visual field) indicated by longer saccades and shorter times to first fixation (Gegenfurtner et al. 2011). Additionally, research on eye-movement modelling examples (EMME) indicates that EMME might help to guide novices' visual attention. EMME illustrates the visual processing behaviour of experts carefully performing a task by recordings their eye movements. A meta-analysis on EMME shows significant effects of eye-tracking measures such as time to first

fixation and fixation duration on novice learners' performance outcomes in terms of learning outcomes and problem-solving (Xie et al. 2021).

Despite the unquestioned potential of the eye-tracking approach, conclusions from eye-tracking data must be drawn very carefully. While there are a variety of eye movement measures, taxonomies, and interpretations, it is important to emphasize the theoretical assumptions as well as the domain and task-specific characteristics that underlie a research objective, and to our knowledge, there is currently no cross-domain taxonomy. However, a categorization based on operational definitions including a variety of eye-tracking measures can be used (see Holmqvist and Andersson 2017; Part III on paradigms and measures). We would like to point out some aspects concerning the ambiguity of eye-tracking measurements. For example, if a participant shows a higher fixation duration for relevant information, this may indicate reasoning, but also confusion or mind-wandering (staring into space). Therefore, eye-tracking should be triangulated with offline and online data channels (which might induce other challenges to a research design), such as self-reports and self-assessments (which often suffer from several biases, Andrade 2019), log files (which are often restricted to binary representations and suffers ambiguous interpretability, Goldhammer et al. 2014), which do not capture relevant off-screen behaviour (Maddox et al. 2018), as well as retrospective (delayed report) or concurrent (disturbing) think aloud (Gegenfurtner and Seppänen 2013), and psychophysiological measures like heart rates which might be hard to interpret (Wu et al. 2014). However, the combination of multimodal data channels within advanced learning technologies is on the rise (Gabriel et al. 2022). Thus, the following review will also take additional behavioural measures into account to underline the potential of multimodal methods in combination with eye-tracking.

Methodology

Identification—Search strategy

A search and selection process in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020; see Liberati et al. 2009; Page et al. 2021) was conducted (Fig. 2). The search was based on common educational research databases (Fraenkel et al. 2019): Web of Science, PsycInfo, and Education Resources Information Center (ERIC). We used the following four major search terms: (1) *eye-tracking AND (2) problem-solving AND (3) training and learning AND (4) computer-based simulation*. Eye-Tracking studies were identified by searching for *eye-tracking OR eye* OR gaze OR fixation OR saccade*. Problem-solving tasks were identified by searching for *problem-solving OR problem* OR decision-making OR decision* OR choice OR domain-specific* OR complex task*. The domains of professional training or education were addressed by *education* OR vocation* OR apprentice* OR training OR program OR workshop OR workplace*. Finally, the search targets computer-based simulations and thus includes *computer OR simulation OR virtual*. The database search yielded 1,061 records. An additional check of reference lists for prominent eye-tracking reviews and meta-analysis was conducted (Gegenfurtner et al. 2011: a meta-analysis on expertise differences in visual comprehension; Lai et al. 2013: a review on eye-tracking studies in learning; Orquin and Mueller Loose 2013, review of attentional shifts in decision tasks). Additionally, Google Scholar was manually searched with varieties of search strings. All findings

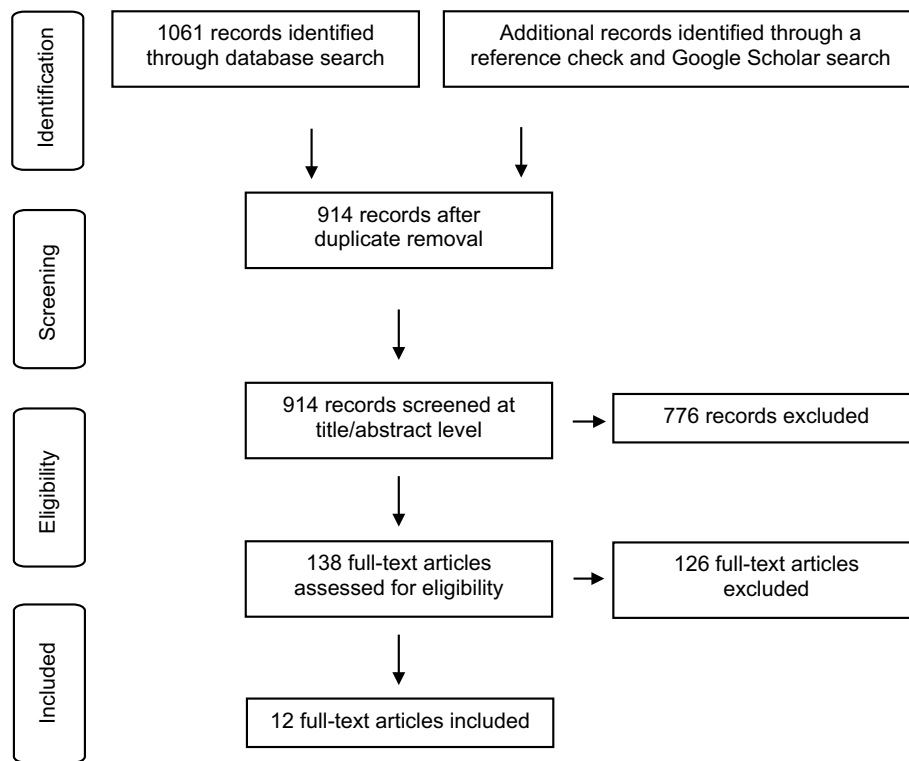


Fig. 2 PRISMA flow chart

Table 1 Key features, inclusion criteria, and exclusion examples

Key features	Inclusion criteria	Exclusion examples
Participants	The study was conducted with healthy human beings as participants	The study was conducted without humans (e.g., animals) or patients with mental or physiological disorders were subjects of interest
Task	The study included a domain-specific problem-solving task	The study included a cross-sectoral or non-authentic problem-solving task (e.g., Tower of Hanoi/London task)
Data collection	The study used eye-tracking as a primary method for data collection	The study did not use eye-tracking as a method (e.g., mentioned eye-tracking for future research)
Type of publication	The article is an empirical study published in a peer-reviewed journal	The article is a review paper, editorial note, or commentary

were combined in a list, and duplicates were identified automatically. An additional manual check was performed. The search process resulted in 914 records for subsequent selection.

Screening—Screening abstracts and titles

This review used CADIMA, a free web tool supporting the systematic review process (see Kohl et al. 2018). The selection criteria were applied at the title and abstract levels. Table 1 lists the key features, inclusion criteria, and exclusion examples.

We tested the inclusion criteria with a consistency check between two independent researchers to examine the inter-reviewer agreement and to overcome researcher bias. A second researcher screened randomly selected abstracts ($n = 45$; > 5% of all potential records). The first check showed that the strength of agreement for screening (either inclusion, exclusion, or unsure) was $k = 0.60$ and considered as moderate (Landis and Koch 1977). A kappa value of at least $k \geq 0.6$ is recommended to continue (Orwin and Vevea 2009; Higgins et al. 2019). Further screening was revised by examining differences and resolving disagreements through discussion and debate by two reviewers, resulting in overall intercoder reliability of $k = 0.89$, which can be interpreted (Landis and Koch 1977) as an almost perfect agreement ($k \geq 0.81$). Given the satisfying kappa value and for the sake of resource efficiency, the process proceeded with a single-person screening. A total of 914 records (after the elimination of duplicates), were screened. Overall, 776 records were excluded from this study.

Eligibility—Screening full texts

No papers were excluded for reasons of inaccessibility. If a paper was not available through institutional access, we kindly asked via ResearchGate® to provide the paper. A total of 138 records remained for screening at the full-text level. Overall, 126 full-text articles were excluded from the study. Owing to the high number of excluded articles, we provide some examples. We excluded studies that (1) did not conduct a complex problem-solving task or (2) were not constructed within an interactive environment.

(1): We excluded studies that did not match a domain-specific problem-oriented approach or interactive design. For example, we excluded a hypermedia-based learning environment that fostered self-regulated learning by using metacognitive pedagogical agents. The subjects learned about the human circulatory system for 90 min. However, a problem-solving task was not in the scope of this study (Taub and Azevedo 2019). Studies based on a fixed stimulus without further interactions (mostly graphs or animations) were excluded. For example, we excluded a multiple-choice science study conducted by Tsai et al. (2011), which examines a science problem represented by four images in a web browser.

(2): We excluded studies on highly interactive learning environments for different reasons. A training simulator often integrates motor-sensitive skills. Training simulators contain realistic switches, knobs, levers, and typical instruments, e.g., flying simulations for pilot training (Schrivier et al. 2008), and offshore drilling simulation (Naqvi et al. 2020). Participants train in dangerous scenarios to become familiar with safety processes in realistic but risk-free simulations and to develop routines. However, the comparability to computer-based learning simulations may not be given, because these training environments also involve motor skills and quick reactions in critical situations. Similarly, we excluded studies that included dummies or actors (O'Meara et al. 2015) and in situ experiments (Esau and Fletcher 2018; Vrzakova et al. 2020). Finally, we excluded papers that were not related to the field of professional training or VET. On this basis, a total of only twelve articles could be included in the analysis.

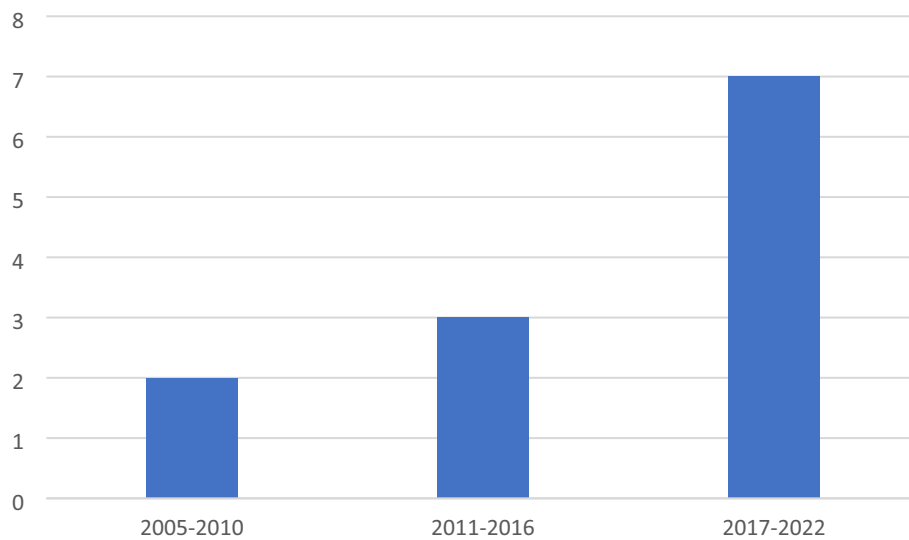


Fig. 3 Included studies per 5-years interval

Table 2 Overview of domains

Domain	Studies
Science	3
Technology	1
Engineering	3
Maths	0
Medical and nursing	3
Air traffic control	1
Military	1

Inclusion—Objectives of analysis and results

Included studies were analysed based on the domain, sample, task, performance measure, eye-tracking devices, measurements calculated, other behavioural measurements collected, analysis techniques, and main findings.

Results

Descriptive

All studies were published between 2005 and 2022, and seven of the studies were published in the last five years (Fig. 3). Most studies were conducted in the United States (6), followed by the Netherlands (3), Germany (1), Israel (1), and Taiwan (1). The sample size ranged between 7 and 70 participants ($M=36.8$, $SD=24.7$). The underlying domains are mainly related to the fields of science, technology, engineering, and mathematics (STEM), especially science (Taub et al. 2017; Emerson et al. 2020; Cloude et al. 2020), engineering (van Gog et al. 2005a; Gomes et al. 2013; Abele 2018), but also the healthcare sector (Lee et al. 2019, 2020; Dubovi 2022) is represented (Table 2).

Usually, students and/or professionals were involved in the studies. Domain-specific complex problem-solving simulations include a broad variety of tasks and findings (see

Appendix Tables 4 and 5 for an overview). Typical tasks include problems like finding the cause of a mysterious disease outbreak (Cloude et al. 2020), fixing malfunctions in electrical circuits (van Gog et al. 2005a), or stabilizing a virtual patient by applying a medical routine (Lee et al. 2019). Performance was most often assessed manually (e.g., by counting the errors solved during troubleshooting), and also by examining pre-post-test scores (e.g., assessing content knowledge on microbiology), or through an automated log file analysis (e.g., by calculating a completion score for the underlying task).

Eye-tracking, measures, and analyses techniques

To answer RQ1, we analysed the different eye-tracking measures that were calculated and the most common analysis techniques that were applied. Further, we examined which additional behavioural measurements were collected next to eye-tracking data. Non-intrusive remote (also screen-based) eye-tracking devices were used in most of the studies (see Table 3 for an overview). Remote setups are typical for experimental lab studies, in contrast to eye-tracking glasses for field studies (e.g., Rosengrant et al. 2021) or virtual reality headsets (e.g., Torres et al. 2017). Currently, webcam-based eye-tracking is examined as a low-cost alternative next to remote eye-tracking devices (Wisiecka et al. 2022).

Following a functionally operational taxonomy for eye-tracking measures (Holmqvist et al. 2011; Holmqvist and Andersson 2017), four types of measures are distinguished in this review: (1) movement measures: including movement measures of direction, amplitudes, duration, velocity, acceleration, shape, sequences and transitions, and scan path comparison measures; *ibid.*, p.439 ff., (2) position measures: including basic positions, measures of position dispersion, similarity, duration, and dilation; *ibid.*, p. 499 ff., (3) count measures: including a variety of countable entities such as saccades, smooth pursuits, blinks, fixations, dwells, AOIs, transitions, and more; *ibid.*, p. 574 ff., as well as (4) latency and distance measures: including the latency of a saccade, pupil dilation latency, eye-mouse distances, and more; *ibid.*, p. 579 ff.

Movement measures refer to the direction of eye movements and scan paths and are infrequently calculated. In total, three studies examined three different movement measures. Gomes et al. (2013) analysed the direction of saccadic movements (next to position measures) to examine eye movement patterns among high and low performers by applying a common unsupervised machine learning clustering (k-means) using trigrams of eye movements. Kang and Landry (2014) conducted a qualitative scan path analysis to examine whether novices in air traffic control follow a professional scan behaviour after being treated with expert scan path examples. Lee et al. (2019) calculated (next to several positions and count measures) the transition rates between AOIs to examine differences among experts and novices in their performance by applying t-test/Mann–Whitney-U and MANOVA.

Position measures are the most common measures used and refer to the positions where participants look. In total, position measures are calculated 15 times (in 11 of the 12 studies). Abele et al. (2017) measured the total fixation duration on relevant AOIs to analyse differences among performance groups by conducting a nonparametric Mann–Whitney-U test. Similar, Sohn et al. (2005) calculated fixation times to determine group differences. Cloude et al. (2020) calculated proportions of total fixation times to predict

Table 3 Overview of eye-tracking setup, calculated measures, analysis techniques, and other collected measures

References	Eye-tracking devices	Eye-tracking measures	Eye-tracking analyses	Add. offline/online measures
Abele et al. 2017	Remote (TobiiX120, 60 Hz)	Total fixation duration (relevant AOs)	Mann–Whitney-U	None
Cloude et al. 2020	Remote (SMI EYERED 250, 30 Hz)	Proportion of total fixation time	Transformation to normal distribution, step-wise simple and multiple linear regression models by using AIC for selection	Log-file of mouse and keyboard data (event and time-based interaction)
Dubovi 2022	Remote (Smart Eye Aurora, 60 Hz)	Time spent (relevant AOs), time to first fixation, fixation counts, fixation dwells, saccade counts, blink rate	ANOVA, linear mixed effects model, prediction (two-phase approach); bivariate correlation with subsequent regression analysis	Facial expression recognition (Affdex), Self-report on affective state (PANAS), EDA (Shimmer 3),
Emerson et al. 2020	Remote (SMI RED 250, 120 Hz)	Proportion of total fixation duration	ANOVA (feature selection), logistic regression (prediction)	Log file data (gameplay behaviour), facial expression recognition (FACET), self-reported Interest-Enjoyment (subscale of Intrinsic Motivation Inventory)
Gomes et al. 2013	Not specified	Number and duration per level on gaze points or fixations, direction of saccadic movement, type of pattern per trigram	k-means clustering	Log-file data (mouse clicks, time per level)
Kang and Landry 2014	Remote (Tobii X60)	Scan path	Scan path analysis according to simple orthogonal shapes	None
Lee et al. 2019	Remote (SMI RED, 250 Hz)	Proportions of dwell time, fixation count, fixation duration, average fixation duration, fixation frequency, transition rate	MANOVA, t-tests, Mann–Whitney U	Log file data (intervention completion score), mental workload (NASA-TLX), systematicity (Hidden Markov model; HMM)
Lee et al. 2020	Remote (SMI RED, 250 Hz)	Pupil diameter, raw gaze points	Linear mixed effects model	Log file data (timestamps for in-game events such as pausing simulation), perceived mental effort
Sohn et al. 2005	Head-mounted (ISCAN ETL-500)	Fixation time	F-test, power law fitting	None
Taub et al. 2017	Remote (SMI EYERED 250, 30 Hz)	Proportion of fixation duration	Multi-level modelling	Log file data (number of books read and reads per book, assessment submission attempts)
Tsai et al. 2016	Remote (faceLAB 4.6, 60 Hz)	Percentage of time spent viewing (PTS), fixation duration (PFD), fixation count (PCD)	Mann–Whitney U, correlations, heatmaps, lag sequential analysis, heat map analysis per group	Self-reported flow experience
(van Gog et al. 2005a)	Remote (SMI, 50 Hz)	Fixation duration, number of fixations	Mann–Whitney U, Friedman test and Nemenyi post-hoc analysis, qualitative matching of verbal and gaze data	Think-aloud (concurrent)

performance differences by applying stepwise simple and multiple linear regression models. Similarly, Emerson et al. (2020) integrated positional gaze data next to students' behavioural traces (such as gameplay behaviour and facial action units) to predict performance and interest groups by running several logistic regression models with different feature compositions. Lee et al. (2020) examined the effects of pausing on the cognitive load within a medical serious game simulation by extracting pupil diameter and applying linear mixed effect models. A multi-level modelling approach was also applied by Taub et al. (2017) including proportions of fixation duration (next to interaction behaviour) to examine differences among performance while accounting for between and within-subject variances (random effects).

Count measures refer to the number and proportions of countable gaze behaviour and are frequently present. Count measures were calculated six times (3 studies) and were most often examined next to similar position measures. Tsai et al. (2016) calculated percentages of fixation counts (next to position measures) to analyse flow experience and visual attention among high and low-performing groups by applying the Mann–Whitney-U test. Also, AOI sequences were used for lag sequential analysis to examine different patterns of visual attention. Van Gog et al. (2005a, b) examined the number of fixations (next to position measures) to examine expertise-related differences by applying Mann–Whitney U, the Friedman test with Nemenyi post-hoc analysis as well as a qualitative data analysis by matching verbal and gaze data.

Finally, latency and distance measures refer to time delays and space distance across eye movements and other points (e.g., mouse cursor). Dubovi (2022) examine latency and distance measures (next to position and count measures) by calculating time to the first fixation and applying ANOVA and linear mixed effect models for group and individual differences as well as regression analysis for performance predictions.

Eye-tracking data can be very ambiguous and is dependent on individual characteristics. To overcome this challenge, researchers are increasingly examining other behavioural data measurements in conjunction with eye-tracking data (Dewan et al. 2019) which can be used for data triangulation. Offline and online measures are frequently used next to eye-tracking (see Table 3). Log files (Gomes et al. 2013; Lee et al. 2019, 2020; Emerson et al. 2020; Cloude et al. 2020) are often collected within computer-based simulations and result in additional and complementary insights into participants' behaviour through mouse clicks and keyboard strokes. Similarly, concurrent think-aloud can help to interpret eye-tracking data through the constant (or retrospective) verbalisation of participants' thoughts (van Gog et al. 2005a). Facial expression recognition (FER) algorithms analyse the expression (Emerson et al. 2020; Dubovi 2022) of anger, disgust, fear, happiness, sadness, surprise, and underlying facial action units (mostly based on the Facial Action Coding System (FACS); Ekman and Friesen 1976). Electrodermal activity (EDA) measures skin conduction as a proxy of psychological or physiological arousal (Dubovi 2022). Additionally, self-report questionnaires are used to measure subjective perceptions (Tsai et al. 2016; Lee et al. 2019, 2020; Emerson et al. 2020; Dubovi 2022).

Main findings

To answer RQ2, we group the main findings related to complex problem-solving in computer-based simulations based on process data measurements.

Most studies analyse differences across performance groups such as high and low performers or expert-novice comparisons and related patterns. Results show strong support for the information-reduction hypothesis following Haider and Frensch (1996; 1999). According to the information reduction hypothesis of Haider and Frensch (1996, 1999), deliberate practice helps learners to ignore redundant information and to focus more on relevant information. Thus, especially experts have learned through practice to distinguish relevant from irrelevant task information (Lee and Anderson 2001) and to use efficient cognitive strategies due to prior experiences (van Merriënboer 2013).

High performers or experts show a longer total fixation time and fewer fixations (Abele et al. 2017), higher proportions of dwell time to total time (with a large effect), a higher ratio of fixation count to total fixation counts (medium effect), and longer fixation duration (large effect) on critically relevant information (Lee et al. 2019). High performers spend more time in a 'problem orientation' and 'action evaluate and next action decision' phase, they spend more fixations on fault-related components, show shorter mean fixation durations in an 'orientation' as well as longer mean fixation durations during a 'formulation' phase (van Gog et al. 2005a, b). It is also reported that through practice, less time is spent on relevant and irrelevant areas (Sohn et al. 2005). Similarly, novices performed better (made fewer false alerts), perceived an expert scan path as useful for their training, and tended to follow a professional expert scan pattern after treatment with an expert scan path (showed a circular movement across the air traffic control screen) (Kang and Landry 2014). A combination of a shorter time to first fixation, fewer clicks, more unique fixations, and longer durations per fixation was found for the high-performance cluster (Gomes et al. 2013), while shorter durations for first fixations might indicate higher attentional readiness and indicates more time spent on reasoning before action. A longer time for the first fixation, a higher number of clicks and short fixation durations might indicate a lack of focus on the strategy or a lack of reasoning (trial and error) before action (Gomes et al. 2013).

In line with the results for experts and high performers, low performers or novices show shorter fixation times, more attention to similar but irrelevant AOIs, and lower visual accuracy. Low performers spend a higher proportion of time gathering information and less time generating hypotheses (Cloude et al. 2020). Low performers or novices show shorter or substantially longer fixation time (a behaviour that might indicate confusion) (Abele et al. 2017). The visual attention was spent on similar medicines, indicating processing difficulties through more fixation counts and dwells (Dubovi 2022). The low comprehension group showed higher mental effort (Tsai et al. 2016) and paid more attention to graphic information (while a high comprehension group spent less on the graphical and more attention on the textual information) examined by qualitative heatmap analysis (Tsai et al. 2016).

Individual differences are related to prior knowledge and differences in cognitive load demands. Lower prior knowledge positively moderated the relation between interaction and fixation on gathering information in a serious game, while a negative relation for higher prior knowledge was found (Cloude et al. 2020). Less successful participants tended to get stuck in messages (cues) and an out-of-screen gaze while successful participants tended to transfer the knowledge and might use an out-of-screen gaze for pausing or reasoning (Tsai et al. 2016) to reduce cognitive load demands. Allowing pauses in a

medical simulation increases the performance and cognitive load, regardless of whether pauses were taken or not. During pauses, the cognitive load was lower than during the simulation. When pauses were available, taking those pauses did not further benefit cognitive load or performance (Lee et al. 2020). Pupillometry might be a valid measure of the cognitive load next to self-reports (Lee et al. 2020).

Other online and offline process measures for behavioural data shed further light on differences in gaze behaviour when solving problems in computer-based simulations. Self-reports showed that a higher flow time distortion was associated with more fixations on the main task while lower flow time distortion was associated with fixations on the message prompts (Tsai et al. 2016). No significant changes in self-reported affective states over time were reported, while a higher level of presence was related to more visual attention to the relevant medicine (Dubovi 2022). Facial expression recognition (FER) shows no significant impact of joy expression on post-tests, but frequent anger expressions were associated with lower post-test scores and positive emotions were related to inducing blinks (Dubovi 2022). EDA shows a significant correlation between EDA peaks and blinks, but not with participants' emotional engagement (Dubovi 2022). Eye-tracking data helps to supplement and contextualize log files (Cloude et al. 2020). Experts show higher levels of systematicity (indicated by the HMM score obtained through a log file analysis) (Lee et al. 2019), and a negative effect between the number of books and performance, as well as for the frequency of books and performance were found. The best performance was associated with reading fewer books but higher frequencies per book, emphasizing a quality reading strategy (fewer books more often) compared to a quantity reading (more books) strategy. Also, no unique association between proportions of fixations on book content or book concept matrix with individual submission attempts were found, but a significant interaction effect between both emphasizes the collection and combination of multichannel data. Low proportions of fixations on book content and concept matrices were related to high performance (Taub et al. 2017). Finally, concurrent think-aloud verbal data for high expertise participants show a predictive behaviour while low expertise participants' verbal data show no orientation and an unstructured initial testing approach.

Behavioural data might be further used for performance prediction within multimodal learning analytics. Gaze as a feature (unimodal) or gameplay, and face as a multimodal feature approach yields an accuracy of .67 for prediction among three performance groups, but adding more modalities comes at the cost of noise, so feature selection must be done carefully to avoid overfitting (Emerson et al. 2020). Also, gameplay and face (multimodal) yield .59 accuracy for prediction among three interest level groups (Emerson et al. 2020). The emotional and cognitive engagement measured via multimodal metrics explained 51% of post-test learning achievements (Dubovi 2022). Interestingly, the blink rate is negatively associated with post-test scores and shows significantly lower rates during the actual problem (Dubovi 2022). Significant associations between performance and the multimodal predictors as well as for the interaction term were found (Taub et al. 2017). The highest performance was related to a higher frequency of books, fewer books, and lower proportions of fixations on book content or concept matrix (Taub et al. 2017). Overall, multimodal data channels are very promising for further progress toward individualized learning analytics approaches (Cloude et al. 2020).

Discussion

This scoping review aimed to analyse the current state of eye-tracking research on domain-specific complex problem-solving in authentic tasks within interactive computer-based simulations. A total of twelve studies from a wide range of vocational education and professional training domains were found.

The most commonly calculated measures are position measures, and these are almost exclusively position duration measures such as the proportion of fixation time or total dwell time. Count measures are also mostly related to the number or proportion of fixations and dwell times. Surprisingly, movement measures are rarely computed and usually refer to saccade directions or scan path. Heatmaps or scan paths are often qualitatively compared. There is a lack of quantitative approaches for measuring time patterns or similarity measurements for a scan path as stated by Holmqvist et al. (2011, 2017). Also, latency and distance measures are almost never calculated. This indicates that the potential to shed further light on complex problem-solving in computer-based simulation might not yet be fully exhausted by calculating other than the standard count and position duration eye-tracking measurements. The much broader variety of potential eye-tracking measures (concerning the underlying specific research questions) should be taken into account (Holmqvist et al. 2011; Holmqvist and Andersson 2017). For example, cognitive load might be measured by considering saccadic peak velocity (Di Stasi et al. 2011), time to first fixations might be an indicator for visual attention to cues and hints in serious games (Conati et al. 2013), and saccade paths (Wu et al. 2014) might be calculated for further insights into behavioural differences and performance predictions.

To analyse eye-tracking data, group comparisons between experts and novices or high-performing and low-performing groups are often computed using common statistical methods such as t-test, (M)ANOVA, or non-parametric Mann–Whitney-U. Patterns between groups are examined with heat maps and lag sequential analyses, by mostly using discrete behaviour codes, or common k-means for clustering purposes. Recently, an increasing number of researchers have focused on individual differences in addition to the group level to account for random effects by applying linear mixed-effects models. This is relevant for eye-tracking research since eye movement data can vary between and within participants over time. The emphasis on the application of mixed-effects models in reading research to analyse eye-tracking data (Catrysse et al. 2018) shows further potential for generalizing results of between-group comparisons while accounting for within-subject variances, and additionally, increases the power of statistical analyses (compared to common approaches such as ANOVA) when conducted for lower aggregated levels (Baayen et al. 2008; Quené and van den Bergh 2008; Catrysse et al. 2018). Finally, post-hoc performance predictions are the first attempts to develop multimodal learning analytics. However, these performance predictions are often performed as subsequent machine learning regression analyses, most of the reported accuracy scores seem to be currently not suited for practical implementation and are currently not integrated within the computer-based simulations for real-time assessments. Research out of the field of decision support systems (Causse et al. 2019) shows promising results to improve performance support.

Similarly, using multimodal data channels seems promising for educational purposes by integrating eye-tracking into systems of multimodal learning analytics

(Cloude et al. 2022). Insights from additional questionnaires, think-aloud protocols, log files, and other psychophysiological measures have proven valuable next to eye-tracking data. Eye-tracking data combined with log file analysis and think-aloud protocols might be useful to validate each other and reveal further information about problem-solving processes (van Gog et al. 2005a, 2005b; Stieff et al. 2011). Interestingly, Taub et al. (2017) found significant effects for the interaction term between gaze and log file data. They emphasize the use of multimodal data by stating “that our most significant results were those that included online trace data from both log files and eye tracking” (Taub et al. 2017, p. 651). In many cases, self-reporting is used as an additional measurement for data triangulation. In addition to eye-tracking, log files and facial expression recognition algorithms are also used. However, few studies use shimmer devices to detect electrodermal activity or practice concurrent thinking aloud. Studies on psychophysiological measures have shown to be valid indicators for problem-solving performance-related measures such as stress (Kärner et al. 2018). Self-reports and log files are useful tools for data triangulation. However, sometimes changes in affective state are not consciously perceived and reported through self-reports but can be measured through facial expression recognition algorithms, as reported by Dubovi (2022). Also, log files show higher systematicity for expert behaviour through HMM scores introduced by Lee et al. (2019). To obtain these systematicity scores, a rigorous task analysis must be performed before computation. Overall, despite the rise of multimodal approaches, the recognition of facial expressions using algorithms, measuring electrodermal activity using shimmer devices, and concurrent (or retrospective, for a comparison see van Gog et al. 2005b) thinking aloud are rare in this sample and data synchronisation remains a challenging aspect of research when data is not collected within a single software, which is not always possible (e.g., when log files of educational data are protected on a separate and secured server).

According to the “information reduction hypothesis” of Haider and Frensch (1996, 1999), deliberate practice helps learners to ignore redundant information and to focus more on relevant information. Thus, experts (and high performers) have learned through practice to distinguish relevant from irrelevant task information (Lee and Anderson 2001) and to use efficient cognitive strategies through experience (van Merriënboer 2013). This is also indicated by many studies in the sample. Similarly, low-performers or novices show shorter fixation times, more attention to similar but irrelevant AOIs, and lower visual accuracy. Performance in computer-based simulations and problem-solving seems to be moderated by prior knowledge, which positively influences the interaction between simulations and information fixation. Lower prior knowledge relates to lower performance and more fiddling around (trial and error strategy). The effects of pausing in simulations (for the medical field) were found to increase performance, whether those pauses were taken or not. Also, successful problem solvers tended to take knowledge from cues when they were given, which could be related to the use of an out-of-screen gaze to pause or think, while unsuccessful participants got stuck in a loop between reading cues and an out-of-screen gaze.

Recently, some work has been done on the post hoc analysis of multimodal features such as eye-tracking and facial expression recognition data as well as log data

for performance prediction. Interestingly, Emerson et al. (2020) state that that using more features for prediction comes at the cost of integrating more noise into the prediction, sometimes making a model's performance worse by overfitting. Some regression-based approaches seem promising and could explain up to 67% of the total variance. Nevertheless, it is difficult to determine machine learning performance by metrics such as accuracy. Especially for the more typical unbalanced datasets, other evaluation metrics such as F1 scores (the harmonic mean of precision and recall) are typically reported in machine learning research.

A major limitation of any literature review is publication bias. By addressing more than one database and a broad search term, as well as additional reference checks, we attempted to challenge publication bias appropriately. Despite these efforts, it is still possible that there is literature available but was not found. Though a review must aim to be all-inclusive, it may not always be possible. A single researcher performed most of the selection procedures. However, acceptable kappa values were calculated for a small number of studies between the two independent coders. Generalisability is not provided over a broad range of domains and tasks. There are shortcomings in the representation of countries and samples. Studies from Western countries are mainly represented within this sample. Also, we want to underline that high performers and experts are not the same (performance-based vs. criteria-based selection). A major shortcoming of this review is the limited number of studies analysed. Thus, we stick to a narrative scoping review but can give no information about the overall statistical effect due to sample restrictions and the heterogeneity in terms of study designs and dependent variables.

Future research might conduct a more systematic review and meta-analysis, particularly on the relationship between performance differences and eye movement measures. So far, within this specific subfield of interest, not enough studies were conducted and published to further examine such relationships. Thus, one advantage of this review is that we show the diversity of eye-tracking as a data collection method as well as different analysis techniques to foster eye-tracking research for VET domains, where computer-based simulations gain increasing relevance for education. This review supports eye-tracking as a data collection method for studying behavioural patterns in learning processes. Further studies should collect experiences during problem-solving processes (Rausch et al. 2019) and learning-related emotions by examining affective states through facial expression recognition (Munshi et al. 2020). Finally, there is a general research gap for eye-tracking studies and behavioural analysis in vocational education and training, and more precisely a vast lack of studies in the field of business education.

Appendix

See Tables 4, 5

Table 4 Descriptive overview of included studies

References	Domain	Sample	Task(s)	Performance measures
Abele et al. 2017	Engineering	15 apprentices in automotive mechatronics engineering from vocational education schools in Germany	Troubleshooting two scenarios in a computer-based motor vehicle simulation	Manual scoring (number of valid errors scored)
Cloude et al. 2020	Microbiology	68 students from three universities in the USA	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game	Pretest–Posttest
Dubovi 2022	Healthcare	61 nursing students from a university in Israel	Collecting clinical data in a virtual hospital and conducting basic medication procedures to solve the complex problem of identifying the correct medication	Pretest–Posttest
Emerson et al. 2020	Microbiology	61 students from three universities in the USA	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game	Pretest–Posttest
Gomes et al. 2013	Engineering	7 high school students from the USA	Solving engineering problems in three short games	Automated Scoring (solve as many levels as possible in a limited time)
Kang and Landry 2014	Air Traffic Control	60 undergraduate and graduate engineering students ($n = 20$ for treatment group) from the USA	Solving possible air traffic conflicts by detecting a loss of separation among multiple aircraft and multiple types of conflicts after training by following an expert scan path	Manual Scoring (number of correct detection/false alarms)
Lee et al. 2019	Medicine	24 professionals and 22 medical students ($n = 46$) from the Netherlands	Stabilising a virtual patient by applying the ABCDE method within a maximum of 15 min in an authentic computer-based medical simulation game	Automated Scoring (intervention completion score by extracting data from log files)
Lee et al. 2020	Medicine	70 medical students from the Netherlands	Stabilising a virtual patient within a maximum of 15 min in an authentic computer-based medical simulation game	Automated Scoring (log files)
Sohn et al. 2005	Military	12 students and staff from the USA	Learning subtasks (identification and control of information search) of tactical decision-making in an anti-air-warfare simulation	Manual Scoring (immediate feedback on timing and accuracy)
Taub et al. 2017	Microbiology	50 non-biology students from the USA	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game	Automated Scoring (log files, in-game assessments as self-regulated judgments of learning)

Table 4 (continued)

References	Domain	Sample	Task(s)	Performance measures
Tsai et al. 2016	Electrophysics	22 university students from Taiwan	Solving a problem-based learning task embedded in a game-based scenario by gathering information and applying knowledge about electromagnets to escape from a lab at the end	Pretest–Posttest
(van Gog et al. 2005a)	Electronics	10 students from the Netherlands	Troubleshooting malfunctions in electrical circuits in a simulation	Manual Scoring (on-task solution aspects)

Table 5 Overview of research questions/hypotheses and main findings

References	Research questions/hypotheses	Main findings
Abele et al. 2017	Successful subjects show a higher total critical fixation duration than less successful subjects	<ul style="list-style-type: none"> - Higher performance longer total fixation time - Lower performance shorter total fixation time or substantial longer fixation time - More prolonged fixation might indicate confusion at some point
Cloude et al. 2020	<p>Time fixating and interacting with scientific reasoning-related game elements predicts post-test scores (RQ1)</p> <p>Time fixating on scientific reasoning-related game elements predicts time interacting with scientific reasoning-related game elements (RQ2)</p> <p>Time fixating on non-scientific reasoning-related game elements predicts the time interacting with non-scientific reasoning-related game elements (RQ3)</p>	<ul style="list-style-type: none"> - Lower performance higher proportion of time gathering information and less time generating hypotheses - Lower prior knowledge positively moderates the relation between interaction and fixation on gathering information in the GBLE while a negative relation for higher prior knowledge was found - No relation between interaction and fixation for non-scientific reasoning was found - Eye-tracking data helps to supplement and contextualize log-files - Multichannel data may be needed for individualized learning analytics approaches
Dubovi 2022	Students' cognitive and emotional engagement can be measured by their self-reports and psycho-physiological real-time measurements and the synergistic effect of cognitive and emotional engagement on learning	<ul style="list-style-type: none"> - Emotional and cognitive engagement via multimodal metrics explained 51% of post-test learning achievement - No significant impact of joy expression on post-test - Frequent anger expressions were associated with lower post-test scores - No significant change in self-reported affective state over three times - Blink rate is negatively associated with post-test scores and shows significantly lower rates during the actual problem - More visual attention is spent on similar medicine, indicating processing difficulties through fixation counts and dwells - Significant correlation between EDA peaks and blinks but not with emotional engagement - Positive emotions were related to inducing blinks - A higher level of presence was related to more visual attention to relevant medicine
Emerson et al. 2020	<p>Student gameplay behavior traces, facial expressions of emotions and eye gaze classify low, medium, and high performing groups. (RQ1)</p> <p>Student gameplay behavior traces, facial expressions of emotions and eye gaze classify low, medium, and high-interest groups. (RQ2)</p>	<ul style="list-style-type: none"> - Gaze as a feature (unimodal) or gameplay + face as a multimodal feature approach yields in the accuracy of 0.67 for prediction among three performance groups - Gameplay + face (multimodal) yields in 0.59 accuracy for prediction among three interest level groups - Adding more modalities comes at the cost of noise, so feature selection must be done carefully to avoid overfitting

Table 5 (continued)

References	Research questions/hypotheses	Main findings
Gomes et al. 2013	Differences in eye-tracking patterns exist between students with low and high performance in the three engineering-related computer games	<ul style="list-style-type: none"> - Shorter time to first fixation, fewer clicks, more unique fixations, and a longer duration per fixation for high-performance cluster - Longer time for the first fixation, a higher number of clicks and short fixation durations might indicate a lack of focus on strategy or reasoning before action ("trial-and-error") - Shorter durations for first fixations might indicate higher attentional readiness and indicates more time spent on reasoning before action
Kang and Landry 2014	The performance will be different for novices exposed to the expert scan path compared to the control group or novices without treatment	<ul style="list-style-type: none"> - Treatment (expert scan path) group showed significantly fewer false alarms than the verbal instruction group or control group - Treated novices tend to follow a professional expert scan pattern after treatment (circular) - Treatment group perceived expert scan paths as helpful, and a scan path could improve the training of novices
Lee et al. 2019	Participants with high domain-specific prior knowledge (DSPK, i.e., experts) show higher systematicity in approach than participants with low DSPK (i.e., novices). (H1) Experts show higher accuracy in visual selection by allocating more visual attention to critical diagnosis areas (H2a) and in motor reactions by completing more interventions (H2b) than novices. Experts show higher speed in performance by completing interventions faster than novices. (H3) Experts experience lower cognitive load than novices. (H4)	<ul style="list-style-type: none"> - Experts-Novice comparison shows for experts: - More systematicity (indicated by HMM score) - Higher proportions of dwell time to total time (large effect), a higher ratio of fixation count to total fixation counts (medium effect), and longer fixation duration (large effect) on critically relevant information - No difference for other AOIs except for the intervention area; a lower proportion of total fixation counts (medium effect) - No difference for average fixation duration and fixation count, but cognitive load and transition rate correlate negatively with self-reported NASA-TLX score
Lee et al. 2020	Cognitive load in the pause-available condition (PA) would be higher than in the pause-unavailable condition (PU), at the overall level. (H1a) Performance in the PA would be higher than in the PU, at the overall level. (H1b) Within PA, cognitive load in the pause-taking group (PAn) would be lower than in the no-pause-taking group (PA0), at the overall level. (H2a) Within the PA, performance in the PAn would be higher than in the PA0, at the overall level. (H2b) In the absence of intense events, the cognitive load would increase during pauses. (H3a) In an intense situation, the cognitive load would decrease during pauses. (H3b)	<ul style="list-style-type: none"> - Overall, allowing pauses increases performance and cognitive load, regardless of whether pauses were taken or not - When pauses were available, taking those pauses did not further benefit cognitive load or performance - During pauses cognitive load was lower compared to simulation - Pupillometry might be a valid measure of the cognitive load next to self-reports
Sohn et al. 2005	Participants learn to pay more attention to task-relevant regions and less attention to task-irrelevant regions with practice over time	<ul style="list-style-type: none"> - Information-seeking behaviour changed over time; reduction in time on relevant and irrelevant regions

Table 5 (continued)

References	Research questions/hypotheses	Main findings
Taub et al. 2017	<p>The more books' participants read, and the more often they read each book, the fewer concept matrix submission attempts they made, resulting in better performance. (H1)</p> <p>The longer fixation durations on the book content and concept matrices, the fewer concept matrix attempts, resulting in better performance. (H2)</p> <p>There will be a significant interaction, such that log file data (number of books and frequency of reading each book) and eye-tracking data (proportions of fixations on book content and book concept matrices) will jointly impact concept matrix submission attempts, with higher levels of all variables resulting in fewer attempts, and thus greater performance. (H3)</p>	<ul style="list-style-type: none"> - Negative effect between number of books and performance as well as for frequency of books and performance - But the best performance was associated with fewer books and higher frequencies per book - Reading more books (quantity) might not improve performance while reading books several times (quality) might do - No unique association between proportions of fixations on book content or book concept matrix with submission attempts were found, but a significant interaction effect. Low proportions of fixations on book content and concept matrices were related to high performance - Significant associations between performance and the multimodal predictors as well as for the interaction term. The highest performance was related to a higher frequency of books, fewer books, and lower proportions of fixations on book content or concept matrix
Tsai et al. 2016	<p>Do players with different conceptual comprehension in GBL:</p> <ul style="list-style-type: none"> - have different visual attention distributions while playing games? If yes, what are the patterns for high and low-achievement players? (RQ1) - have different patterns of visual attention transactions (representing the players' control strategies of multi-tasking coordination applied in the game)? (RQ2) - experience different levels of game flow? (RQ3) 	<p>Low comprehension group:</p> <ul style="list-style-type: none"> - Higher PFD and PCD in the components area - Higher mental effort (heatmap) - Viewed graphical information more frequently than the high comprehension group - Paid more attention to graphic information according to heatmap analyses (while the high comprehension group spent less attention on graphical and more attention on textual information) - Low comprehension group tended to get stuck in the message (cues) and out-of-screen gaze while the successful group tend to transfer knowledge and might use out-of-screen gaze as a pausing/reasoning strategy <p>high comprehension group showed</p> <ul style="list-style-type: none"> - A higher sense of control and concentration, <p>flow experience and visual attention association:</p> <p>higher flow time distortion fixations on the main task</p> <p>lower flow time distortion fixations on the message prompt</p>

Table 5 (continued)

References	Research questions/hypotheses	Main findings
(van Gog et al. 2005a)	Higher expertise participants spend more time on problem orientation, problem formulation, deciding on actions and evaluating them, while lower expertise participants are more likely to test out the functioning of the circuit to try to generate new hypotheses. (1) Higher expertise participants' orientation and evaluation phase will be less cognitively demanding than reasoning, and all these processes are more demanding for lower than higher expertise participants. In the 'problem orientation' phase, higher expertise participants will have a higher proportion of fixations on components related to major faults. (2) Eye movement and concurrent verbal protocol data together show how eye movement data may make to the investigation of cognitive processes. (3)	- High expertise participants spent more time in the 'problem orientation' and 'action evaluate & next action decision' phase (but not for the 'problem formulation' phase) - Higher expertise groups only differ for mean fixation duration over all phases, but show more fixations on fault-related components, they show: Shorter mean fixation duration in the 'orientation' phase Longer mean fixation duration in the 'problem formulation' phase Verbal data that reveals predictive behaviour - Low expertise participants' verbal data show no orientation and an unstructured initial testing approach

Acknowledgements

We would like to thank Georg Dariush Gorshid for his initial assistance in the study selection.

Author contributions

CM conceived the aim of the review and coordinated the review process. The design of the review, analysis of the review of the literature and the writing of the draft was performed by CM. CM, JS, and AR wrote, reviewed, and edited the manuscript in several rounds. All authors read and approved the final manuscript.

Funding

Not applicable.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Declarations**Competing interests**

The authors declare that they have no competing interests.

Received: 14 October 2022 Accepted: 25 March 2023

Published online: 18 April 2023

References

- Abele S (2018) Diagnostic problem-solving process in professional contexts: theory and empirical investigation in the context of car mechatronics using computer-generated log-files. *Vocat Learn* 11:133–159. <https://doi.org/10.1007/s12186-017-9183-x>
- Abele S, Ostertag R, Peissner M, Schuller A (2017) Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess. Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg [An eye tracking study on the problem-solving process in professional contexts: relevance of "representing information" for the diagnostic problem-solving success]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik* 113:86–109
- Alemdag E, Cagiltay K (2018) A systematic review of eye tracking research on multimedia learning. *Comput Educ* 125:413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Anderson JR, Bothell D, Douglass S (2004) Eye movements do not reflect retrieval processes: limits of the eye-mind hypothesis. *Psychol Sci* 15:225–231
- Andrade HL (2019) A critical review of research on student self-assessment. *Front Educ* 4:87. <https://doi.org/10.3389/educ.2019.00087>
- Arksey H, O'Malley L (2005) Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 8:19–32. <https://doi.org/10.1080/1364557032000119616>
- Ashraf H, Sodergren MH, Merali N, Mylonas G, Singh H, Darzi A (2018) Eye-tracking technology in medical education: a systematic review. *Med Teach* 40:62–69. <https://doi.org/10.1080/0142159X.2017.1391373>

- Baayen RH, Davidson DJ, Bates DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *J Mem Lang* 59:390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Beck K, Landenberger M, Oser F (eds) (2016) Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based competence measurement in vocational education and training: results from the BMBF funding initiative ASCOT]. wbv media, Bielefeld
- Breckwoldt J, Gruber H, Wittmann A (2014) Simulation learning. In: Billett S, Harteis C, Gruber H (eds) *International handbook of research in professional and practice-based learning*. Springer Netherlands, Dordrecht, pp 673–698
- Brehmer B, Dörner D (1993) Experiments with computer-simulated microworlds: escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Comput Hum Behav* 9:171–184. [https://doi.org/10.1016/0747-5632\(93\)90005-D](https://doi.org/10.1016/0747-5632(93)90005-D)
- Catrysse L, Gijbels D, Donche V, De Maeyer S, Lesterhuis M, Van den Bossche P (2018) How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *Br J Educ Psychol* 88:118–137. <https://doi.org/10.1111/bjep.12181>
- Causse M, Lancelot F, Maillant J, Behren J, Cousy M, Schneider N (2019) Encoding decisions and expertise in the operator's eyes: using eye-tracking as input for system adaptation. *Int J Hum Comput Stud* 125:55–65. <https://doi.org/10.1016/j.ijhcs.2018.12.010>
- Clarebout G, Elen J, Lowyck J, Van den Ende J, Van den Enden E (2009) Tropical medicine open learning environment. In: Rogers PL, Berg GA, Boettcher JV, Howard C, Justice L, Schenk KD (eds) *Encyclopedia of distance learning*, 2nd edn. IGI Global, Hershey, pp 2155–2159
- Cloude EB, Dever DA, Wiedbusch MD, Azevedo R (2020) Quantifying scientific thinking using multichannel data with crystal island: implications for individualized game-learning analytics. *Front Educ* 5:572546. <https://doi.org/10.3389/educ.2020.572546>
- Cloude EB, Azevedo R, Winne PH, Biswas G, Jang EE (2022) System design for using multimodal trace data in modeling self-regulated learning. *Front Educ* 7:928632. <https://doi.org/10.3389/educ.2022.928632>
- Conati C, Jaques N, Muir M (2013) Understanding attention to adaptive hints in educational games: an eye-tracking study. *Int J Artif Intell Educ* 23:136–161. <https://doi.org/10.1007/s40593-013-0002-8>
- Dewan MAA, Murshed M, Lin F (2019) Engagement detection in online learning: a review. *Smart Learn Environ* 6:1. <https://doi.org/10.1186/s40561-018-0080-z>
- Di Stasi L, Antoli A, Canas J (2011) Main sequence: an index for detecting mental workload variation in complex tasks. *Appl Ergon* 42:807–813. <https://doi.org/10.1016/j.apergo.2011.01.003>
- Dörner D (1987) *Problemlösen als Informationsverarbeitung [Problem solving as information processing]*. Kohlhammer, Stuttgart
- Dörner D (1997) *The logic of failure: recognizing and avoiding error in complex situations*. Basic Books, New York
- Dörner D, Funke J (2017) Complex problem solving: what it is and what it is not. *Front Psychol* 8:1153. <https://doi.org/10.3389/fpsyg.2017.01153>
- Dubovi I (2022) Cognitive and emotional engagement while learning with VR: the perspective of multimodal methodology. *Comput Educ* 183:104495. <https://doi.org/10.1016/j.compedu.2022.104495>
- Duchowski AT (2017) *Eye tracking methodology*. Springer International Publishing, Cham
- Duncker K (1945) On problem-solving. *Psychol Monogr* 58:i–113. <https://doi.org/10.1037/h0093599>
- Ekman P, Friesen WV (1976) Measuring facial movement. *Environ Psychol Nonverbal Behav* 1:56–75
- Emerson A, Cloude EB, Azevedo R, Lester J (2020) Multimodal learning analytics for game-based learning. *Br J Educ Technol* 51:1505–1526. <https://doi.org/10.1111/bjet.12992>
- Esau T, Fletcher S (2018) Prozessorientierte Analyse von konstruktiven Problemlöseprozessen auf Basis von Eye-Tracking-Aufnahmen [Process-oriented analysis of engineering-design problem solving processes based on the eye-tracking recording]. *J Technol Educ* 6:2198–306. <https://doi.org/10.48513/joted.v6i1.116>
- Fischer A, Neubert JC (2015) The multiple faces of complex problems: a model of problem solving competency and its implications for training and assessment. *J Dyn Decis Mak* 1:6–6. <https://doi.org/10.11588/jddm.2015.1.23945>
- Fraenkel JR, Wallen NE, Hyun HH (2019) *How to design and evaluate research in education*. McGraw Hill, New York
- Frensch PA, Funke J (1995) Definitions, traditions, and a general framework for understanding complex problem solving. In: Frensch PA, Funke J (eds) *Complex problem solving: the European perspective*. Lawrence Erlbaum, Hillsdale, pp 3–25
- Funke J (2012) Complex problem solving. In: Seel NM (ed) *Encyclopedia of the sciences of learning*. Springer, Boston, pp 682–685
- Gabriel F, Cloude EB, Azevedo R (2022) Using learning analytics to measure motivational and affective processes during self-regulated learning with advanced learning technologies. In: “Elle” Wang Y, Joksimović S, San Pedro MOZ, Way JD, Whitmer J (eds) *Social and emotional learning and complex skills assessment*. Springer International Publishing, Cham, pp 93–108
- Gegenfurtner A, Seppänen M (2013) Transfer of expertise: an eye tracking and think aloud study using dynamic medical visualizations. *Comput Educ* 63:393–403. <https://doi.org/10.1016/j.compedu.2012.12.021>
- Gegenfurtner A, Lehtinen E, Säljö R (2011) Expertise differences in the comprehension of visualizations: a meta-analysis of eye-tracking research in professional domains. *Educ Psychol Rev* 23:523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Goldhammer F, Naumann J, Stelter A, Tóth K, Rölke H, Klieme E (2014) The time on task effect in reading and problem solving is moderated by task difficulty and skill: insights from a computer-based large-scale assessment. *J Educ Psychol* 106:608–626. <https://doi.org/10.1037/a0034716>
- Gomes JS, Yassine M, Worsley M, Blikstein P (2013) Analysing Engineering Expertise of High School Students Using Eye Tracking and Multimodal Learning Analytics. In: D’Mello S, Calvo R, Olney A (eds) *Proceedings of the 6th International Conference on Educational Data Mining*, Memphis, p 3
- Haider H, Frensch PA (1996) The role of information reduction in skill acquisition. *Cogn Psychol* 30:304–337

- Haider H, Frensch PA (1999) Eye movement during skill acquisition: more evidence for the information-reduction hypothesis. *J Exp Psychol Learn Mem Cogn* 25:172–190. <https://doi.org/10.1037/0278-7393.25.1.172>
- Hannafin MJ (1995) Open-ended learning environments: foundations, assumptions, and implications for automated design. In: Tennyson RD, Barron AE (eds) *Automating instructional design: computer-based development and delivery tools*. Springer, Berlin, Heidelberg, pp 101–129
- Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (2019) *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons, Hoboken
- Holmqvist K, Andersson R (2017) *Eye tracking: a comprehensive guide to methods, paradigms, and measures*, 2nd edn. CreateSpace, Charleston
- Holmqvist K, Nyström M, Andersson R, Dewhurst R, Jarodzka H, Van de Weijer J (2011) *Eye tracking: a comprehensive guide to methods and measures*. OUP Oxford, Oxford
- Jarodzka H (2021) Research Methods in Multimedia Learning. In: Mayer RE, Fiorella L (eds) *The Cambridge handbook of multimedia learning*, 3rd edn. Cambridge University Press, Cambridge, pp 41–54
- Jarodzka H, Holmqvist K, Gruber H (2017) Eye tracking in educational science: theoretical frameworks and research agendas. *J Eye Mov Res* 10:1–18. <https://doi.org/10.16910/JEMR.10.1.3>
- Jonassen DH (2000) Toward a design theory of problem solving. *Education Tech Research Dev* 48:63–85. <https://doi.org/10.1007/BF02300500>
- Jonassen DH, Hung W (2008) All problems are not equal: implications for problem-based learning. *Interdiscip J Problem-Based Learn*. <https://doi.org/10.7771/1541-5015.1080>
- Just MA, Carpenter PA (1980) A theory of reading: from eye fixations to comprehension. *Psychol Rev* 87:329. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kang Z, Landry SJ (2014) Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Hum Factors* 56:1150–1162. <https://doi.org/10.1177/0018720814523066>
- Kärner T, Minkley N, Rausch A, Schley T, Sembill D (2018) Stress and resources in vocational problem solving. *Vocat Learn* 11:365–398. <https://doi.org/10.1007/s12186-017-9193-8>
- Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, Wilhelm R (2018) Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. *Environ Evid* 7:8. <https://doi.org/10.1186/s13750-018-0115-5>
- Lai M-L, Tsai M-J, Yang F-Y, Hsu C-Y, Liu T-C, Lee SW-Y, Lee M-H, Chiou G-L, Liang J-C, Tsai C-C (2013) A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educ Res Rev* 10:90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>
- Lajoie SP, Naismith L (2012) Computer-based learning environments. In: Seel NM (ed) *Encyclopedia of the sciences of learning*. Springer, Boston, pp 716–718
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174. <https://doi.org/10.2307/2529310>
- Lee FJ, Anderson JR (2001) Does learning a complex task have to be complex? A study in learning decomposition. *Cogn Psychol* 42:267–316
- Lee JY, Donkers J, Jarodzka H, van Merriënboer J (2019) How prior knowledge affects problem-solving performance in a medical simulation game: using game-logs and eye-tracking. *Comput Hum Behav* 99:268–277. <https://doi.org/10.1016/j.chb.2019.05.035>
- Lee JY, Donkers J, Jarodzka H, Sellenraad G, van Merriënboer J (2020) Different effects of pausing on cognitive load in a medical simulation game. *Comput Hum Behav* 110:106385. <https://doi.org/10.1016/j.chb.2020.106385>
- Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ* 339:b2700. <https://doi.org/10.1136/bmj.b2700>
- Maddox B, Bayliss AP, Fleming P, Engelhardt PE, Edwards SG, Borgonovi F (2018) Observing response processes with eye tracking in international large-scale assessments: evidence from the OECD PIAAC assessment. *Eur J Psychol Educ* 33:543–558. <https://doi.org/10.1007/s10212-018-0380-2>
- Mayer RE (1998) Cognitive, metacognitive, and motivational aspects of problem solving. *Instr Sci* 26:49–63. <https://doi.org/10.1023/A:1003088013286>
- Mayer RE, Wittrock MC (2006) Problem solving. In: Alexander PA, Winne PH (eds) *Handbook of educational psychology*, 2nd edn. Routledge, pp 287–303
- Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E (2018) Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 18:143. <https://doi.org/10.1186/s12874-018-0611-x>
- Munshi A, Mishra S, Zhang N, Paquette L, Ocumpaugh J, Baker R, Biswas G (2020) Modeling the relationships between basic and achievement emotions in computer-based learning environments. In: Bittencourt II, Cukurova M, Muldner K, Luckin R, Millán E (eds) *Artificial intelligence in education*. Springer International Publishing, Cham, pp 411–422
- Naqvi S, Raza M, Ghazal S, Salehi S, Kang Z, Teodoru C (2020) Simulation-based training to enhance process safety in offshore energy operations: process tracing through eye-tracking. *Process Saf Environ Prot* 138:220–235. <https://doi.org/10.1016/j.psep.2020.03.016>
- O'Meara P, Munro G, Williams B, Cooper S, Bogossian F, Ross L, Sparkes L, Browning M, McClounan M (2015) Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: a proof of concept paper. *Int Emerg Nurs* 23:94–99. <https://doi.org/10.1016/j.ienj.2014.11.001>
- Orquin JL, Mueller Loose S (2013) Attention and choice: a review on eye movements in decision making. *Acta Psychol* 144:190–206. <https://doi.org/10.1016/j.actpsy.2013.06.003>

- Orwin RG, Vevea JL (2009) Evaluating coding decisions. In: Cooper H, Hedges LV, Valentine JC (eds) *The handbook of research synthesis and meta-analysis*, vol 2. Russell Sage Foundation, New York, pp 177–203
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. <https://doi.org/10.1136/bmj.n71>
- Quené H, van den Bergh H (2008) Examples of mixed-effects modeling with crossed random effects and with binomial data. *J Mem Lang* 59:413–425. <https://doi.org/10.1016/j.jml.2008.02.002>
- Rausch A, Seifried J, Wuttke E, Kögler K, Brandt S (2016) Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empiri Res Vocat Educ Train* 8:9. <https://doi.org/10.1186/s40461-016-0035-y>
- Rausch A, Kögler K, Seifried J (2019) Validation of embedded experience sampling (EES) for measuring non-cognitive facets of problem-solving competence in scenario-based assessments. *Front Psychol* 10:1200. <https://doi.org/10.3389/fpsyg.2019.01200>
- Rayner K (1998) Eye movements in reading and information processing: 20 years of research. *Psychol Bull* 124:372
- Rosengrant D, Herrington D, O'Brien J (2021) Investigating student sustained attention in a guided inquiry lecture course using an eye tracker. *Educ Psychol Rev* 33:11–26. <https://doi.org/10.1007/s10648-020-09540-2>
- Schrivver AT, Morrow DG, Wickens CD, Talleur DA (2008) Expertise differences in attentional strategies related to pilot decision making. *Hum Factors* 50:864–878. <https://doi.org/10.1518/001872008X374974>
- Sohn M, Douglass S, Chen M, Anderson J (2005) Characteristics of fluent skills in a complex, dynamic problem-solving task. *Hum Factors* 47:742–752. <https://doi.org/10.1518/001872005775570943>
- Stieff M, Hegarty M, Deslongchamps G (2011) Identifying representational competence with multi-representational displays. *Cogn Instr* 29:123–145. <https://doi.org/10.1080/07370008.2010.507318>
- Strohmaier AR, Schiepe-Tiska A, Chang Y-P, Müller F, Lin F-L, Reiss KM (2020) Comparing eye movements during mathematical word problem solving in Chinese and German. *ZDM Int J Math Educ* 52:45–58. <https://doi.org/10.1007/s11858-019-01080-6>
- Taub M, Azevedo R (2019) How does prior knowledge influence eye fixations and sequences of cognitive and metacognitive SRL processes during learning with an intelligent tutoring system? *Int J Artif Intell Educ* 29:1–28. <https://doi.org/10.1007/s40593-018-0165-4>
- Taub M, Mudrick NV, Azevedo R, Millar GC, Rowe J, Lester J (2017) Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with *Crystal Island*. *Comput Hum Behav* 76:641–655. <https://doi.org/10.1016/j.chb.2017.01.038>
- Theeuwes J (2010) Top-down and bottom-up control of visual selection. *Acta Physiol* 135:77–99. <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Torres F, Neira Tovar LA, del Rio MS (2017) A learning evaluation for an immersive virtual laboratory for technical training applied into a welding workshop. *Eurasia J Math Sci Technol Educ* 13:521–532
- Tsai M-J, Hou H-T, Lai M-L, Liu W-Y, Yang F-Y (2011) Visual attention for solving multiple-choice science problem: an eye-tracking analysis. *Comput Educ* 58:375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>
- Tsai M-J, Huang L-J, Hou H-T, Hsu C-Y, Chiou G-L (2016) Visual behavior, flow and achievement in game-based learning. *Comput Educ* 98:115–129. <https://doi.org/10.1016/j.compedu.2016.03.011>
- Underwood G, Everatt J (1992) The role of eye movements in reading: some limitations of the eye-mind assumption. In: Chekaluk E, Llewellyn K (eds) *Advances in psychology*. Elsevier, Amsterdam, pp 111–169
- van Merriënboer JJG (2013) Perspectives on problem solving and instruction. *Comput Educ* 64:153–160. <https://doi.org/10.1016/j.compedu.2012.11.025>
- van Gog T, Jarodzka H (2013) Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In: Azevedo R, Aleven V (eds) *International handbook of metacognition and learning technologies*. Springer, New York, pp 143–156
- van Gog T, Paas F, van Merriënboer JJG (2005a) Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data: uncovering expertise-related differences. *Appl Cognit Psychol* 19:205–221. <https://doi.org/10.1002/acp.1112>
- van Gog T, Paas F, van Merriënboer JJG, Witte P (2005b) Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *J Exp Psychol Appl* 11:237–244. <https://doi.org/10.1037/1076-898X.11.4.237>
- van Gog T, Jarodzka H, Scheiter K, Gerjets P, Paas F (2009) Attention guidance during example study via the model's eye movements. *Comput Hum Behav* 25:785–791. <https://doi.org/10.1016/j.chb.2009.02.007>
- Van Ostaeyen S, Embo M, Schellens T, Valcke M (2022) Training to support ePortfolio users during clinical placements: a scoping review. *MedSciEduc* 32:921–928. <https://doi.org/10.1007/s40670-022-01583-0>
- Vrzakova H, Begel A, Mehtätalo L, Bednarik R (2020) Affect recognition in code review: an in-situ biometric study of reviewer's affect. *J Syst Softw* 159:110434
- Weinert FE (2001) Concept of competence: a conceptual clarification. In: Rychen DS, Salganik LH (eds) *Defining and selecting key competencies*. Hogrefe & Huber Publishers, Cambridge, pp 45–65
- Williamson DM, Mislavy RJ, Bejar II (2006) Automated scoring of complex tasks in computer-based testing: an Introduction. In: Williamson DM, Mislavy RJ, Bejar II (eds) *Automated scoring of complex tasks in computer-based testing*. Psychology Press, London, pp 1–13
- Wisiecka K, Krejtz K, Krejtz I, Sromek D, Cellary A, Lewandowska B, Duchowski A (2022) Comparison of webcam and remote eye tracking. In: 2022 Symposium on eye tracking research and applications. ACM, Seattle WA USA, pp 1–7

- Wu C-H, Tzeng Y-L, Huang YM (2014) Understanding the relationship between physiological signals and digital game-based learning outcome. *J Comput Educ* 1:81–97. <https://doi.org/10.1007/s40692-014-0006-x>
- Xie H, Zhao T, Deng S, Peng J, Wang F, Zhou Z (2021) Using eye movement modelling examples to guide visual attention and foster cognitive performance: a meta-analysis. *J Comput Assist Learn* 37:1194–1206. <https://doi.org/10.1111/jcal.12568>
- Yang F-Y, Tsai M-J, Chiou G-L, Lee SW-Y, Chang C-C, Chen L-L (2018) Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies. *J Educ Technol Soc* 21:28–45

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
