Educ Psychol Rev (2011) 23:523–552
DOI 10.1007/s10648-011-9174-7

REVIEW

# Expertise Differences in the Comprehension of Visualizations: a Meta-Analysis of Eye-Tracking Research in Professional Domains

**Andreas Gegenfurtner · Erno Lehtinen · Roger Säljö**

**Abstract** This meta-analysis integrates 296 effect sizes reported in eye-tracking research on expertise differences in the comprehension of visualizations. Three theories were evaluated: Ericsson and Kintsch's (Psychol Rev 102:211–245, 1995) theory of long-term working memory, Haider and Frensch's (J Exp Psychol Learn Mem Cognit 25:172–190, 1999) information-reduction hypothesis, and the holistic model of image perception of Kundel *et al*. (Radiology 242:396–402, 2007). Eye movement and performance data were cumulated from 819 experts, 187 intermediates, and 893 novices. In support of the evaluated theories, experts, when compared with non-experts, had shorter fixation durations, more fixations on task-relevant areas, and fewer fixations on task-redundant areas; experts also had longer saccades and shorter times to first fixate relevant information, owing to superiority in parafoveal processing and selective attention allocation. Eye movements, reaction time, and performance accuracy were moderated by characteristics of visualization (dynamics, realism, dimensionality, modality, and text annotation), task (complexity, time-on-task, and task control), and domain (sports, medicine, transportation, other). These findings are discussed in terms of their implications for theories of visual expertise in professional domains and their significance for the design of learning environments.

**Keywords** Eye tracking · Expertise · Graphics comprehension · Long-term working memory · Information reduction · Parafoveal processing · Meta-analysis

Expertise in the comprehension of visualizations has gained growing attention over the past years (de Groot and Gobet 1996; Ericsson and Lehmann 1996; Haider and Frensch 1999;

---

A. Gegenfurtner (✉) · E. Lehtinen · R. Säljö
Centre for Learning Research and Department of Teacher Education, University of Turku, Assistentinkatu 7, 20014 Turku, Finland
e-mail: angege@utu.fi

R. Säljö
Department of Education, Communication, and Learning, University of Gothenburg, Läroverksgatan 15, 40530 Gothenburg, Sweden

Kalyuga 2007; Mann *et al*. 2007; Vickers 2007; Hyönä 2010; Krupinski 2010), largely because of the fascinating observations about how experts are able to solve complex tasks after glancing very briefly at a picture. For example, expert radiologists are able to detect cancer in a mammogram in a split second (Kundel *et al*. 2007), and to take a very different example, grandmasters locate the positions of checking pieces on the chessboard without moving the eye (Reingold *et al*. 2001). Eye-tracking methodology has provided significant insight into some of the perceptual mechanisms underlying expert performance in a range of professional settings, including aviation (Schriver *et al*. 2008), fish classification (Jarodzka *et al*. 2010), car driving (Crundall *et al*. 1999), arts (Vogt and Magnussen 2007), and sports (North *et al*. 2009). In the present study, expert performance is understood as "consistently superior performance on a specified set of representative tasks for a domain" (Ericsson and Lehmann 1996, p. 277).

One problem with eye-tracking methodology in expertise research is the typically small sample size. Individual study findings are therefore likely to be influenced by sampling error. This may explain some of the disagreements in the literature. In particular, some authors reported that experts have more fixations of shorter duration than novices (Konstantopoulos 2009; Litchfield *et al*. 2008). Others reported findings in the opposite direction (Bertrand and Thullier 2009; Vogt and Magnussen 2007). To account for effect size heterogeneity, the present study uses meta-analytic methods to cumulate individual research findings after controlling for sampling error; the goal was to evaluate three theories that account for expert superiority in visual domains. A second purpose was to estimate the extent to which the quality of visualizations and task characteristics moderate the size of the expertise differences.

## Theories Explaining the Reproducibility of Expert Superiority in Visual Domains

Expertise in the comprehension of visualizations can be explained by several theories. First, the theory of long-term working memory (Ericsson and Kintsch 1995) focuses on qualitative changes in memory structures. This theory assumes that expertise extends the capacities for information processing owing to the acquisition of retrieval structures that allow advanced learners to rapidly encode information in long-term memory and efficiently access it for later task operations. The theory of long-term working memory proposes that the limited-capacity assumption of working memory (Cowan 2001; Miller 1956) needs to be reconsidered when experts perceive domain-specific visual material.[1] If we assume that eye movements reflect the processes underlying task performance (Just and Carpenter 1984) and if it is true that experts encode and retrieve information more rapidly than novices, then it follows that experts' rapid information processing should be reflected in shorter fixation durations.

Expertise in the comprehension of visualizations can be explained by a second theory. The information-reduction hypothesis (Haider and Frensch 1999) focuses on the learned selectivity of information processing. This theory proposes that expertise optimizes the

---

[1] Early research on chess masters (Chase and Simon 1973; Jongman 1968 as cited in de Groot and Gobet 1996, Chapter 4) already indicated that extended memory capacities allow experts to recognize more and larger perceptual chunks than novices. These researchers also introduced the idea that chunking at least partly involves recognition ("revisualization"; Jongman 1968) of visual material in long-term memory. Perceptual chunking now constitutes an important perceptual learning mechanism (cf. the concept of *unitization*, Goldstone 1998).

amount of processed information by neglecting task-irrelevant information and actively focusing on task-relevant information, which is accomplished through strategic considerations to selectively allocate attentional resources. Haider and Frensch (1999, p. 188) noted that, because of learning and training, "redundant information is perceptually ignored whenever this is possible." Information reduction thus results from a growing ability to differentiate between the variables of a stimulus array (Gibson 1986). If the assumption is true that experts select the amount of information they attend to, then it follows that experts should have fewer fixations of shorter duration on task-redundant areas and more fixations of longer duration on task-relevant areas.

Finally, expertise in the comprehension of visualizations can be explained by a third theory. The holistic model of image perception (Kundel *et al.* 2007) focuses on the extension of the visual span. This theory proposes that expertise changes the temporal organization of perceptual processes in such a way that they allow advanced learners to proceed from an initial global analysis toward a finer-grained decomposition into hierarchical structural components: Experts are assumed to extract information from widely distanced and parafoveal regions (see also Charness *et al.* 2001; Reingold *et al.* 2001). It follows that to discern between signal and noise, experts do not need to bring information into the fovea. If expertise extends the visual span through parafoveal processing, then the ability for holistic analysis should be reflected in longer saccade length and in shorter times to first fixate task-relevant areas.

Assumptions and operationalizations of these three theories are shown in Table 1 and definitions of the operationalizations in Table 2. While each of these three theories addresses different aspects, they are, of course, not mutually exclusive: The theories provide complementary accounts of some of the mechanisms underlying the reproducibility of expert task superiority when comprehending domain-specific visualizations. Each theory can be generalized across a range of visualization and task characteristics. In particular, Ericsson and Kintsch (1995) illustrated the theoretical value of long-term working memory with a series of examples of skilled performance in the professions. In addition, assumptions of the information-reduction hypothesis were corroborated with different stimuli, including textual (Haider and Frensch 1999) and pictorial (Jarodzka *et al.* 2010) material. However, all these studies had a relatively small sample size. Moreover, this research used differing study settings. In addition to the above theories, it seems important to pay attention to how the size of expertise differences in performance and eye movements can vary. Variance in the size of the difference may be moderated by characteristics of visualization and task.

**Table 1** Assumptions and operationalizations of the three theories of visual expertise

| Theory | Assumption | Operationalization |
|---|---|---|
| Theory of long-term working memory | Rapid information processing through acquisition of retrieval structure | Fixation duration |
| Information-reduction hypothesis | Selective attention allocation | Number of fixations (relevant),<br>number of fixations (redundant),<br>fixation duration (relevant),<br>fixation duration (redundant) |
| Holistic model of image perception | Global analysis through extended visual span and parafoveal processing | Time to first fixate (relevant),<br>saccade length |

**Table 2** Definition of eye movement and performance variables included in the meta-analysis

| Variable | Definition |
|---|---|
| Fixation | Miniature eye movements that relatively stabilize the retina for a prolonged posture of the eyes over an object |
| Number of fixations (relevant) | Number of fixations on areas with information relevant for task completion |
| Number of fixations (redundant) | Number of fixations on areas with information irrelevant for task completion |
| Fixation duration | Time of 1 fixation |
| Fixation duration (relevant) | Time of 1 fixation on areas with information relevant for task completion |
| Fixation duration (redundant) | Time of 1 fixation on areas with information irrelevant for task completion |
| Time to first fixate (relevant) | Time between stimulus onset and the first fixation on an area with information relevant for task completion |
| Saccade length | Distance of eye movements between 2 fixations |
| Response time | Time between stimulus onset and response |
| Performance accuracy | Ratio between correct responses and all responses |

For extended discussions on eye movement physiology, see Gibson (1986, Chapter 12), Just and Carpenter (1984), and Vickers (2007, Chapter 1)

## Visualizations as a Moderator of the Size of Expertise Differences

Characteristics of visualizations[2] systematically moderate the size of expertise differences. To explain the moderating effect, it is useful to review theories of multimedia learning. A large body of evidence into cognitive load theory (Paas *et al*. 2003; Sweller 1994; Sweller *et al*. 2011) and the cognitive theory of multimedia learning (Mayer 2009; Mayer and Moreno 2003) suggest that visualizations are easier to comprehend when they are designed in accordance with human cognitive architecture. The gap between experts and novices may be narrowed in several ways.

*Reducing extraneous processing demands* First, the gap can be narrowed by reducing extraneous processing demands. Some material is difficult to process because it poses a high extraneous load on limited working memory capacity. Extraneous material is information that is not needed to comprehend visualizations, which makes this material redundant (Mayer 2009; Sweller 1994). Because experts have acquired retrieval cues that extend their working memory capacity for domain-specific material (Ericsson and Kintsch 1995) and because experts tend to perceptually ignore redundant information (Haider and Frensch 1999), we assume that extraneous material is particularly detrimental for novices. Such extraneous material is more often found in realistic than in schematic visualizations. Per definition, schematic visualizations reduce the number of cues and depict only relevant information in abstract, simplified form. Several studies found that schematic visualizations are superior for learning in comparison to realistic visualizations (Dwyer and Joseph 1984; Scheiter *et al*. 2009). Moreover, visuospatial demands on working memory are more

---

[2] By visualization, we mean all kinds of pictorial representations of information, including (but not limited to) pictures, graphs, diagrams, concept maps, video films, animations, and simulations.

strongly reduced in static than in dynamic visualizations, at least if both materials represent comparable information, are non-interactive, and are not presented in user-paced segments (Scheiter *et al.* 2009; Mayer 2009; Spanjers *et al.* 2010). In continuous dynamic visualizations, the holding of information already attended to and the processing of new incoming information increase the amount of information held in working memory and so conflict with the limited resources of processing capacity available (Barrouillet and Camos 2007; Mayer and Moreno 2003). Analogously, the same capacity limitations apply when working with three-dimensional material. Navigating in two-dimensional visualizations requires less representational holding over time as new information is processed, so two-dimensional rather than three-dimensional material—under comparable conditions—is assumed to reduce extraneous processing demands for novices' working memory.

*Fostering generative processing of essential material* The gap between experts and novices can be narrowed in a second way. Some visualizations are easier to comprehend because they contain elements that foster the generative processing of essential material. For example, some visualizations present related sources of information that could not be understood in isolation, so they use both visual and verbal channels to deliver information (cf. modality principle; Mayer 2009; Moreno and Mayer 1999). In a recent meta-analysis of 43 effect sizes, Ginns (2005, p. 327) concluded that multimodal representations manage limited processing resources in working memory and so "improve the chances these resources are committed by students to schema construction and automation, rather than processes extraneous to learning." Another way to foster generative processing is to annotate visual material by text. The rationale for combining words and pictures is intrinsic in the multimedia principle (Mayer 2009). Numerous studies have indicated the relevance of the multimedia principle when people are unfamiliar with a domain. For example, in a classic study, Mayer and Gallini (1990) showed that students with low prior knowledge scored highest on conceptual recall and problem solving in conditions that combined text with illustrations. Recent findings support the benefit of combining visual and textual material to foster mental model development (Butcher 2006) and posttest retention (Eilam and Poyas 2008). To summarize, these studies indicate that the multimedia principle seems to foster the processing of essential material when people have low domain-specific knowledge.

*Inducing an expertise reversal effect* The gap between experts and novices can be narrowed in a third way: by delimiting the performance of experts. Some of the principles of visualization design that are beneficial for novices are detrimental for experts (for a recent review on the expertise reversal effect, see Kalyuga 2007). For example, when learners have high prior knowledge or high visuospatial ability, the multimedia principle seems to lose its relevance. In that case, people learn more from pictures alone than from a combination of pictures with words. In particular, Plass *et al.* (2003) found that high-spatial ability learners had more correct answers in a German vocabulary test in a visual-only treatment than in a verbal-only or combined verbal + visual treatment. Similarly, a diagram-only format was most effective for apprentices who had received prior training in machine drilling (Kalyuga *et al.* 2000). Expertise reversal effects also seem to exist for the level of dynamics, at least when the visualization is schematic and two-dimensional; specifically, in graph transformation tasks, less knowledgeable learners benefited more from static than from animated examples, while more knowledgeable learners benefited more from animated rather than from static examples (Kalyuga 2008). In short, these findings indicate how expertise can alter instructional efficiency and how different characteristics of a visualization moderate the size of expert–novice differences. Although differences are widely documented for performance data, differences are rarely investigated for eye movement data.

In summary, because they were shown to reduce extraneous processing demands (Barrouillet and Camos 2007; Scheiter *et al.* 2009), foster generative processing of essential material (Ginns 2005; Moreno and Mayer 1999), and partly induce expertise reversal effects (Kalyuga 2007; Kalyuga *et al.* 2000), we assumed that expertise differences in eye movement characteristics were smaller for static, schematic, or two-dimensional visualizations that used dual modality or text annotation.

## Task as a Moderator of the Size of Expertise Differences

Besides visualization characteristics, characteristics of the task are assumed to moderate the size of expertise differences. In particular, we focus on three aspects: task complexity, time-on-task, and task control. First, tasks differ as a function of their contextual demands. In a recent meta-analysis of sport expertise, Mann *et al.* (2007) proposed that expertise differences in the perceptual strategies and decision-making strategies of athletes are task dependent. It seems reasonable to assume that this statement has validity also to domains beyond sports. Based on Wood's (1986) general theoretical model of tasks and Campbell's (1986) complex task classification, four levels of task complexity in eye-tracking research on the comprehension of visualizations can be discerned: viewing tasks, detection tasks, decision tasks, and problem-solving tasks (see Table 3 for an overview). An example of a *viewing task* is reported by Vickers (1988, study 1, p. 55), who instructed gymnasts to "watch each sequence of slides very carefully." An example of a *detection task* can be found in Moreno *et al.* (2006, p. 864), who instructed swimming coaches "to detect as many errors as possible" in the performance of a swimmer. An example of a *decision task* is reported in Vaeyens *et al.* (2007), who instructed soccer players facing simulations of offensive patterns of play to make a tactical decision on the next action out of three available options. Finally, a *problem-solving task* is exemplified by Van Gog *et al.* (2005), who asked participants to troubleshoot malfunctioning electrical circuits so that they would function again properly. These task examples suggest that contextual demands differ as a function of the number of desired outcomes, the multiplicity of paths to attain desired outcomes, and the coordinative complexity of informational cues in the task material while

**Table 3** Four levels of task complexity in the comprehension of visualizations

| Task type | Multiplicity of paths | Number of desired outcomes | Coordinative complexity | Example |
|---|---|---|---|---|
| Viewing task | Low | Low | Low | Watching paintings or video films |
| Detection task | Low | Low | High | Detecting errors or target objects |
| Decision task | Low | High | High | Deciding between a set of given options |
| Problem-solving task | High | High | High | Troubleshooting or generating solutions |

moving toward task completion (Campbell 1986; Wood 1986). It can therefore be expected that because task complexity varies across studies, the differences in eye movement patterns and task performance can vary.

Time-on-task is another characteristic that can moderate expertise differences. Time-on-task can be unlimited or limited. When time is unlimited, evidence suggests that experts tend to spend less time-on-task completion than novices (Mann *et al.* 2007) because of experts' superior speed in information processing (Ericsson and Kintsch 1995; Haider and Frensch 1999) and higher levels of confidence (Nodine *et al.* 2002). In particular, according to perceptual control theory (Powers 2005), human actions are driven by perceived differences between the internal state of a situation (desired outcome, e.g., correctly diagnosing a medical case) and the external state of a situation (final medical diagnosis). While approaching external states, novices tend to show lower levels of confidence, so they allocate more temporal resources for re-examining the task procedure. In contrast, when time is limited, the experimental control sets a cutoff value for task completion, which delimits variation in temporal resource allocation. For this reason and the theoretical premises noted above, we expect that expertise differences vary as a function of time limitation, being more articulated in unlimited time-on-task settings.

Task control is a third characteristic that can moderate expertise differences. While some task settings are non-interactive, other task settings afford human–computer interaction, which offer participants a certain degree of freedom to navigate in the visualization (Wilson *et al.* 2010). Segmentation of dynamic visualizations can also afford user control (Spanjers *et al.* 2010). It is generally assumed that non-experts perform better when a visualization is user-paced rather than system-paced because user control allows the user to regulate visuospatial processing demands in working memory (cf. segmenting principle; Mayer 2009). Hence, we expect that expertise differences were moderated by task control, being smallest for user-controlled tasks.

In summary, differences in eye movements and performance may vary as a function of task complexity, time-on-task, and task control. Following Mann *et al.* (2007), we also assumed that expertise differences might vary simply by the fact that they are situated in different domains. We therefore add in our meta-analysis a moderator variable that tracks the domain characteristic to account for heterogeneity in effect sizes.

## The Present Study—Hypotheses

The present study focuses on expertise-related differences in the comprehension of visualizations. The independent variable was expertise. The dependent variables were the number of fixations, number of fixations on relevant areas, number of fixations on redundant areas, fixation duration, fixation duration on relevant areas, fixation duration on redundant areas, time to first fixate relevant areas, saccade length, response time, and performance accuracy. All dependent variables are listed and defined in Table 2.

One aim of the present study was to cumulate observed effect sizes in eye-tracking research by correcting the variance across studies for the bias of sampling error. Hypotheses were based on three theories. Based on the theory of long-term working memory (Ericsson and Kintsch 1995), experts were assumed to show shorter fixation durations than novices (hypothesis 1). Based on the information-reduction hypothesis (Haider and Frensch 1999), experts were expected to have fewer fixations of shorter duration on task-redundant areas (hypothesis 2a) and more fixations of longer duration on task-relevant areas (hypothesis 2b). Based on the holistic model of image perception (Kundel *et al.* 2007), experts were hypothesized to have longer saccade length (hypothesis 3a) and shorter times to first fixate task-relevant areas (hypothesis 3b).

A second aim of the present study was to estimate the extent to which the hypothesized moderator variables may explain the remaining variance in results. Moderators were hypothesized at three levels: visualization, task, and domain. At the visualization level, smaller differences were assumed when the visualization was static (hypothesis 4a), schematic (hypothesis 4b), two-dimensional (hypothesis 4c), when it used a dual modality (hypothesis 4d), and when the visualization was annotated by text (hypothesis 4e). At the task level, smaller differences were expected for less complex tasks (hypothesis 5a) performed with a time limit (hypothesis 5b) and for tasks controlled by the user (hypothesis 5c). Finally, at the domain level, we hypothesized differences as a function of the professional domain (hypothesis 6).

## Method

### Literature searches and criteria for inclusion

Meta-analysis (Hunter and Schmidt 2004) was used to estimate expertise-related differences in the comprehension of visualizations. Studies were located that reported group differences between experts and novices, experts and intermediates, or intermediates and novices. To be included in the database, a study had to report a point biserial correlation $r_{\text{pb}}$, mean and standard deviation for each group, or other effect sizes that could be converted to $r_{\text{pb}}$ (Cohen's $d$; $F$, $t$, $\chi^2$, or $Z$ statistics). Formulae reported in Rosenthal and DiMatteo (2001) were used to convert effect sizes. The database includes studies that report data from healthy samples. Studies reporting therapeutic data were excluded. Comparisons of melodic fragments in musical sight-reading were also omitted because the reading of music notation in staves, scores, or tablatures resembles processes of text comprehension. We included studies using a quasi-experimental contrastive approach of expert performance while we excluded longitudinal studies of novice training, as training reports typically indicate relative levels of performance improvement, not expert performance itself. We also omitted studies on word or scene recognition because these stimuli represent verbal descriptive symbols or natural events rather than graphic representations of these symbols and events. Finally, the database also includes studies that report the comprehension of visualizations in simulations and in online, virtual, or computer-mediated settings.

Using these inclusion criteria, studies published up to December 2010 were located in several ways. First, we searched the PsycINFO, PubMed, and Web of Science databases using relevant keywords included in the titles or abstracts of English-language journals. Keywords were (1) for expertise: *expert*, *novice*, *skilled*, *elite*, and *expertise* and (2) for eye movements: *eye tracking*, *fixations*, *saccades*, and *eye movements*. This search revealed a preliminary 248 articles: 140 from PsycINFO, 90 from PubMed, and 18 from Web of Sciences. Elimination of duplicates revealed a sample of 194 articles. Of these, 21 articles met all inclusion criteria. Second, we searched the Web of Science and Google Scholar databases to cross-reference these articles as well as recent reviews of expert performance (Ericsson *et al.* 2006; Krupinski 2010; Mann *et al.* 2007), which resulted in an additional 44 articles meeting the criteria.

### Coding of variables

From the database, a total of 65 articles, book chapters, conference papers, and dissertations were eventually categorized as codable because they contributed at least one effect size to the meta-analysis. Uncertainty in judging codability was resolved through discussion among all authors. Two independent raters coded a randomly selected subset, approximately 15.39% of the studies from the final sample (ten publications). Because intercoder reliability was generally high (Cohen's $\kappa$=0.92), one rater continued to code the remaining studies. The included studies are preceded by an asterisk in the reference list and summarized in the "Appendix". Because one of the aims of this meta-analysis was to identify moderator variables that can account for effect size heterogeneity, different characteristics were tabulated from the research literature. Specifically, each study was coded for effect size estimates of expertise differences as well as for characteristics of visualization, task, and domain.

*Effect size estimates of expertise differences* Differences between experts and novices, experts and intermediates, and intermediates and novices were tracked at two levels: perception and performance. At the perception level, we coded the difference in the number of fixations on task-relevant, task-irrelevant, and all areas; duration of fixations on task-relevant, task-irrelevant, and all areas; time to first fixate on task-relevant areas; and saccade length. At the performance level, we coded the difference in performance accuracy and reaction time. In addition, we coded the first author, publication year, the number of participants in each expertise group, their age (in years), gender (percentage of females), and experience (in years).

*Visualization characteristics* Coded visualization characteristics included the level of dynamics (static, dynamic), realism (schematic, rather realistic, photo-realistic), dimension (two-dimensional, three-dimensional), modality (visual only, visual plus auditory), and text annotation (not annotated, annotated).

*Task characteristics* Coded task characteristics included the level of task complexity (viewing task, detection task, decision task, problem-solving task), time-on-task (limited, unlimited), and task control (system-control, user-control).

*Domain characteristics* In addition to task characteristics, we also coded the professional domain, including sports (team sports, e.g., soccer, baseball; one-on-one sports, e.g., chess, boxing; and solo sports, e.g., gymnastics, swimming), medicine (e.g., radiology, cardiology, laparoscopic surgery), transportation (aviation, car driving), and other (e.g., cartography, forensics, physics).

## Computation and analysis of effect sizes

Analysis occurred in two stages. A primary meta-analysis aimed to compute the corrected effect size estimate of the expertise difference in all perception and performance variables. A meta-analytic moderator analysis then aimed to identify moderator effects in those effect size estimates. Both analyses are specified in turn.

The primary meta-analysis was done using the methods of meta-analysis of correlations corrected individually for artifacts (Hunter and Schmidt 2004). First, the difference between experts and novices, experts and intermediates, and intermediates and novices was quantified and converted to the effect size index $r_{pb}$ (Rosenthal and DiMatteo 2001). Next, the distribution of $r_{pb}$ was corrected for sampling error to get $r_c$. Note that the correction was done using frequency-weighted average, not Fisher's $z$ transformation, since the latter has been shown to produce upwardly biased correlation estimates (Hall and Brannick 2002). Finally, standard deviations of the corrected observed correlation were calculated to compute 99% confidence intervals around $r_c$. Point biserial correlations range from −1.00 to +1.00. A negative value indicates that experts had, for example, a smaller number of fixations or shorter fixation durations than novices. Analogously, a positive value indicates that experts had a higher number of fixations or longer fixation durations than novices.

The meta-analytic moderator analysis followed the primary meta-analysis. We estimated the a priori hypothesized moderators using theory-driven sub-group analyses. A critique on using sub-groups is that it reduces the number of data sources per analysis, resulting in second-order sampling error. Although this study contained a large number of data sources and participants, the possibility of second-order sampling error cannot be completely ruled out. It is therefore indicated when warranted for interpreting the results.

Table 6 presents the results of the primary meta-analysis, and Table 7 presents the results of the meta-analytic moderator analysis. Each cell in Tables 6 and 7 represents one individual meta-analysis; thus, the two tables are the culmination of a series of over 120 individual meta-analyses. Inputs into the meta-analyses include effect size estimates in the form of mean observed correlations, along with the number of data sources and sample sizes.

Some authors advocate performing an outlier analysis. Because current approaches for identifying outlier coefficients in meta-analytic data sets tend to over-identify small correlations relative to large ones (Beal *et al*. 2002), the removal of outlying cases becomes problematic. Furthermore, the formula for sampling error variance used in the present study allows and corrects for occasional extreme outlying values. It follows that eliminating outliers can overcorrect for sampling error and underestimate $\mathrm{SD}_{rc}$ (Hunter and Schmidt 2004). Based on these problems with outlier removal in meta-analytic work, calculations in the present study were based on the full data set. All calculations were based on the assumption that the population parameter values $\rho$ for expertise-related perception and performance differences vary from study to study, so we used a random-effects model to obtain more accurate estimates of the width of the confidence intervals (National Research Council 1992; Schmidt *et al*. 2009). In addition, if a study reported more than one effect size, a single composite variable was created to comply with the assumption of independence. As an exception to this rule, linear composites were not created for the theoretically predicted moderator variables, since composite correlations would have obscured moderator effects and prohibited further analysis. Multiple experiments reported in one study were treated as independent data sources. For this reason and the statistical decisions noted above, the 65 located articles reported 73 independent data sources with 296 effect sizes, which we included in our analysis.

## Results

### Description of the included studies

Table 4 presents the mean sample size of experts, intermediates, and novices. On average, expertise research using eye-tracking methodology draws on a sample of 11 experts, 10 intermediates, and 12 novices. The small number of participants in each group signals the presence of sampling error, which seems to justify a meta-analysis for cumulating individual study findings. Total sample size of the meta-analysis was 819 experts, 187 intermediates, and 893 novices. Table 5 presents a summary of the number of data sources, participants, and participant characteristics. Overall, 53 data sources (72.60%) examined differences between experts and novices, 3 data sources (4.17%) examined differences

**Table 4** Sample size of experts, intermediates, and novices by professional domain

| | Experts | | Intermediates | | Novices | |
|---|---|---|---|---|---|---|
| Professional domain | *M* | SD | *M* | SD | *M* | SD |
| Sports | 11.17 | 5.70 | 10.70 | 5.12 | 11.37 | 5.65 |
| Team sports | 12.33 | 5.22 | 16.67 | 4.62 | 12.00 | 4.01 |
| One-on-one sports | 11.27 | 6.31 | 8.40 | 2.61 | 12.50 | 7.19 |
| Solo sports | 6.00 | 2.92 | 7.50 | 3.54 | 5.80 | 3.11 |
| Medicine | 7.73 | 6.87 | 5.60 | 2.61 | 8.36 | 7.39 |
| Transportation | 16.58 | 6.45 | 17.33 | 2.31 | 23.33 | 21.76 |
| Other | 8.56 | 3.36 | – | – | 9.89 | 3.79 |
| Total | 11.22 | 6.30 | 10.39 | 5.61 | 12.76 | 11.26 |

**Table 5** Number of data sources, participants, and participant characteristics by professional domain

| | | | Age | | Gender | | Experience | |
|---|---|---|---|---|---|---|---|---|
| Professional domain | *k* | *N* | *M* | SD | *M* | SD | *M* | SD |
| Experts | | | | | | | | |
| Sports | 41 | 458 | 23.96 | 4.65 | 15.32 | 29.17 | 10.87 | 2.87 |
| Team sports | 25 | 297 | 23.34 | 3.35 | 1.92 | 8.13 | 10.75 | 2.48 |
| One-on-one sports | 11 | 131 | 24.28 | 2.02 | 26.23 | 23.73 | 11.90 | 0.00 |
| Solo sports | 5 | 30 | 26.88 | 11.79 | 66.67 | 57.74 | 10.73 | 5.59 |
| Medicine | 11 | 85 | – | – | – | – | 11.77 | 8.25 |
| Transportation | 12 | 199 | 31.50 | 11.08 | 27.54 | 23.03 | 12.49 | 9.46 |
| Other | 9 | 77 | 32.48 | 4.84 | 59.52 | 13.53 | 7.02 | 5.25 |
| Total | 73 | 819 | 26.54 | 7.38 | 21.18 | 28.81 | 11.04 | 5.56 |
| Intermediates | | | | | | | | |
| Sports | 10 | 107 | 21.39 | 4.73 | 30.11 | 45.39 | 3.41 | 3.61 |
| Team sports | 6 | 80 | 22.79 | 3.40 | 10.23 | 20.46 | 2.85 | 4.94 |
| One-on-one sports | 2 | 12 | 23.80 | 0.00 | 0.00 | 0.00 | – | – |
| Solo sports | 2 | 15 | 14.80 | 3.39 | 100.0 | 0.00 | 4.23 | 1.05 |
| Medicine | 5 | 28 | – | – | – | – | 4.77 | 6.69 |
| Transportation | 3 | 52 | 31.15 | 10.31 | 20.00 | 34.64 | 12.67 | 10.79 |
| Other | 0 | – | – | – | – | – | – | – |
| Total | 18 | 187 | 23.65 | 7.27 | 27.36 | 41.28 | 6.07 | 7.30 |
| Novices | | | | | | | | |
| Sports | 38 | 432 | 22.98 | 3.62 | 21.22 | 33.37 | 3.88 | 3.50 |
| Team sports | 22 | 272 | 22.42 | 3.27 | 3.47 | 13.89 | 4.37 | 3.97 |
| One-on-one sports | 11 | 131 | 24.94 | 2.79 | 29.70 | 25.56 | 3.80 | 0.00 |
| Solo sports | 5 | 29 | 21.83 | 5.93 | 93.33 | 11.55 | 1.88 | 1.70 |
| Medicine | 11 | 92 | – | – | – | – | 0.15 | 0.35 |
| Transportation | 12 | 280 | 20.41 | 1.52 | 32.93 | 23.17 | 0.74 | 1.25 |
| Other | 9 | 89 | 26.77 | 4.39 | 40.56 | 13.58 | 0.14 | 0.30 |
| Total | 70 | 893 | 22.99 | 3.95 | 26.20 | 30.19 | 2.21 | 3.12 |

between experts and intermediates, and 15 data sources examined differences between experts, intermediates, and novices (20.83%). Expert participants reported about a decade of experience in their domain (cf. 10-year rule); at the same time, similar amounts of experience are also reported for transportation intermediates. Across professional groups, experts were older than both intermediates [$t$ (48)=3.28, 99% confidence interval (CI)=0.41; 5.31] and novices [$t$ (12)=3.56, 99% CI=0.65; 6.49]. Age difference between the intermediate and novice sample was marginal, as were all gender differences.

## Eye movement and performance differences

Table 6 summarizes the results of the primary meta-analyses on eye movement and performance differences between experts and novices, experts and intermediates, and intermediates and novices. For each meta-analysis, (1) the columns are the number of data

**Table 6** Psychometric properties of eye movement and performance differences

| Variable | $k$ | $N$ | $r$ | $r_c$ | $SD_{rc}$ | 99% CI |
|---|---|---|---|---|---|---|
| Expert–novice | | | | | | |
| Number of fixations | 43 | 949 | −0.06 | −0.04 | 0.41 | −0.07; −0.01 |
| Number of fixations (relevant) | 8 | 185 | 0.56 | 0.53 | 0.22 | 0.49; 0.57 |
| Number of fixations (redundant) | 3 | 65 | −0.32 | −0.31 | 0.13 | −0.35; −0.27 |
| Fixation duration | 44 | 1,165 | −0.05 | −0.09 | 0.28 | −0.11; −0.07 |
| Fixation duration (relevant) | 15 | 325 | 0.29 | 0.27 | 0.41 | 0.21; 0.33 |
| Fixation duration (redundant) | 8 | 147 | −0.49 | −0.43 | 0.23 | −0.48; −0.39 |
| Time to first fixate (relevant) | 7 | 125 | −0.40 | −0.31 | 0.12 | −0.34; −0.28 |
| Saccade length | 8 | 196 | 0.29 | 0.30 | 0.27 | 0.25; 0.35 |
| Response time | 37 | 1,050 | −0.45 | −0.38 | 0.29 | −0.40; −0.36 |
| Performance accuracy | 46 | 1,175 | 0.49 | 0.45 | 0.27 | 0.43; 0.47 |
| Expert–intermediate | | | | | | |
| Number of fixations | 9 | 206 | −0.27 | −0.25 | 0.34 | −0.31; −0.19 |
| Fixation duration | 10 | 267 | −0.01 | 0.00 | 0.18 | −0.03; 0.03 |
| Fixation duration (relevant) | 3 | 62 | 0.13 | 0.07 | 0.34 | −0.04; 0.19 |
| Fixation duration (redundant) | 2 | 44 | −0.19 | −0.01 | 0.25 | −0.11; 0.09 |
| Time to first fixate (relevant) | 3 | 35 | −0.28 | −0.23 | 0.30 | −0.37; −0.09 |
| Saccade length | 5 | 91 | 0.36 | 0.38 | 0.15 | 0.34; 0.42 |
| Response time | 8 | 213 | −0.45 | −0.41 | 0.19 | −0.44; −0.38 |
| Performance accuracy | 7 | 204 | 0.45 | 0.45 | 0.22 | 0.41; 0.49 |
| Intermediate–novice | | | | | | |
| Number of fixations | 6 | 125 | −0.17 | −0.23 | 0.36 | −0.31; −0.15 |
| Fixation duration | 6 | 144 | 0.05 | 0.04 | 0.16 | 0.01; 0.07 |
| Fixation duration (relevant) | 2 | 52 | 0.01 | 0.03 | 0.09 | 0.00; 0.06 |
| Fixation duration (redundant) | 2 | 44 | −0.33 | −0.26 | 0.12 | −0.31; −0.21 |
| Time to first fixate (relevant) | 3 | 37 | −0.04 | 0.01 | 0.27 | −0.11; 0.13 |
| Saccade length | 3 | 43 | −0.03 | −0.12 | 0.20 | −0.20; −0.04 |
| Response time | 5 | 130 | −0.46 | −0.43 | 0.39 | −0.52; −0.34 |
| Performance accuracy | 4 | 113 | 0.48 | 0.44 | 0.14 | 0.41; .0.48 |

sources $k$, (2) the total sample size $N$, (3) the uncorrected mean observed correlation $r$, (4) the corrected mean observed correlation $r_c$, (5) the standard deviation of $r_c$, and (6) the 99% CI around $r_c$. Because not all studies entered in the analysis—though having manipulated the group variable—were intervention studies, the psychometric properties shown in Table 6 should be interpreted as indicating the magnitude of the difference, not a causal effect of expertise on eye movement and performance variables. As regards performance, results in Table 3 show that experts had shorter reaction times and higher performance accuracy than novices; similar trends exist for expert–intermediate and intermediate–novice comparisons. As regards eye movements, results are presented in relation to the theory of long-term working memory, the information-reduction hypothesis, and the holistic model of image perception.

*Theory of long-term working memory* Consistent with the theory of long-term working memory, experts had slightly shorter fixation durations than novices ($r=-0.09$). There was no difference between experts and intermediates ($r=0.00$). Unexpectedly, intermediates had slightly longer fixation durations than novices ($r=0.04$). This result was based on a small cell size, so we shall note that the latter finding may be due to second-order sampling error. Still, the finding may also indicate nonmonotone aspects of expertise development (Lesgold *et al.* 1988).

*Information-reduction hypothesis* Experts and novices showed a similar number of fixations, with experts fixating slightly less ($r=-0.04$). Consistent with the assumptions of the information-reduction hypothesis, the expert sample had more fixations on task-relevant areas ($r=0.53$) and fewer fixations on task-redundant areas ($r=-0.31$) than the novice sample had. The number of fixations on total areas was smaller for experts than for intermediates ($r=-0.25$) and smaller for intermediates than for novices ($r=-0.23$).

Differences in the fixation duration on relevant and redundant areas showed moderate to strong effect sizes for expert–novice comparisons ($r_{\text{relevant}}=0.27$, $r_{\text{redundant}}=-0.43$). Generally, as expertise levels increase, there was a tendency for longer fixation durations on relevant areas ($r_{\text{expert–intermediate}}=0.07$, $r_{\text{intermediate–novice}}=0.03$) and for shorter fixation durations on redundant areas ($r_{\text{expert–intermediate}}=-0.01$, $r_{\text{intermediate–novice}}=-0.26$).

*Holistic model of image perception* Experts had shorter times to first fixate task-relevant areas than had intermediates ($r=-0.23$) or novices ($r=-0.30$). Experts also had longer saccadic amplitudes than intermediates ($r=0.38$) or novices ($r=0.31$). These findings are in line with assumptions of the holistic model of image perception. Unexpectedly, the intermediate sample had marginally longer times to first fixate task-relevant areas ($r=0.01$) and a slightly shorter saccade length ($r=-0.12$) than the novice sample.

## Moderator effects

High standard deviations in the primary meta-analyses indicate effect size heterogeneity, so estimation of effects of the hypothesized moderator variables seems warranted. To minimize the bias of second-order sampling error, we focused on those variables with both the highest number of data sources $k$ and the highest sample size $N$. These were the number of fixations, fixation duration, reaction time, and performance accuracy for the expert–novice comparison. Moderator effects were hypothesized for visualization, task, and domain characteristics.

*Visualization characteristics* Table 7 clearly shows that eye movement and performance variables differ as a function of the visualization characteristics. On dynamics, experts tended to employ more fixations ($r$=0.05) of shorter duration ($r$=−0.12) for dynamic visualizations; novices tended to employ more fixations ($r$=−22) of shorter duration ($r$= 0.04) for static visualizations. In terms of realism, experts gradually used more fixations than novices did as the levels of realism increased, from schematic ($r$=−0.24) to rather realistic ($r$=0.01) and photo-realistic ($r$=0.02) visualizations. No clear pattern was identified for fixation duration. Expert–novice differences were smaller for photo-realistic than for schematic visualizations regarding response time ($r_{\text{photo-realistic}}$=−0.36 vs. $r_{\text{schematic}}$=−0.48) and performance accuracy ($r_{\text{photo-realistic}}$=−0.44 vs. $r_{\text{schematic}}$=−0.47). On dimensionality, experts tended to employ fewer fixations for two-dimensional ($r$=−0.07) and more fixations for three-dimensional ($r$=0.27) visualizations than novices; for both dimension levels, experts had shorter fixation durations, shorter reaction times, and a higher performance accuracy. On modality and text annotation, expert–novice differences in reaction time were smaller for annotated visualizations ($r_{\text{annotated}}$=−0.28 vs. $r_{\text{not annotated}}$=−0.42) in dual modality ($r_{\text{visual}}$=−0.45 vs. $r_{\text{visual plus auditory}}$=−0.31). Novices were more accurate when visualizations were annotated by text (0.35) than when not (0.48).

*Task characteristics* Task complexity moderated expert–novice differences. The difference in performance accuracy increased with growing levels of complexity ($r_{\text{detection}}$=0.33 vs. $r_{\text{decision}}$=0.46 vs. $r_{\text{problem-solving}}$=0.58). Across all complexity levels, experts had shorter response times than novices; no clear pattern emerged for the number of fixations and fixation duration. Concerning time-on-task and task control, experts had fewer fixations ($r_{\text{system}}$=−0.04 and $r_{\text{unlimited}}$=−0.36) of shorter duration ($r_{\text{system}}$=−0.09 and $r_{\text{unlimited}}$=−0.04) for system-controlled tasks of unlimited time. Unexpectedly, experts were more rapid and more accurate in task completion than novices were when the task was user-controlled (0.56 vs. 0.43).

*Domain characteristics* Expert–novice differences varied as a function of the professional group. No clear pattern emerged for eye movement parameter. Differences in reaction time were generally highest for sport domains.

## Discussion

One aim of this meta-analysis has been to cumulate research on expertise that used eye-tracking methodology by correcting the variance across studies for the bias of sampling error. A second aim has been to estimate the moderating effects of visualization, task, and domain characteristics. The heterogeneity of and the disagreements in the literature ultimately led this study to seek a better understanding of whether, to what extent, and under which conditions experts, intermediates, and novices differed in the comprehension of visualizations.

### Discussion of primary meta-analyses

The primary meta-analyses aimed at testing the assumptions of three theories. Table 8 shows the cumulative evidence. First, results confirmed assumptions of the theory of long-term working memory (Ericsson and Kintsch 1995; hypothesis 1) with regard to experts and novices. Unexpectedly, however, intermediates had longer durations than novices did.

**Table 7** Expert–novice differences as a function of visualization, task, and domain characteristics

| | Number of fixations | | | | Fixation duration | | | | Response time | | | | Performance accuracy | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moderator variable | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD |
| Visualization characteristics | | | | | | | | | | | | | | | | |
| Dynamics | | | | | | | | | | | | | | | | |
| Static | 14 | 313 | −0.22 | 0.45 | 11 | 226 | 0.04 | 0.33 | 11 | 234 | −0.47 | 0.31 | 13 | 276 | 0.43 | 0.23 |
| Dynamic | 29 | 636 | 0.05 | 0.36 | 33 | 939 | −0.12 | 0.26 | 26 | 816 | −0.35 | 0.28 | 33 | 899 | 0.45 | 0.29 |
| Realism | | | | | | | | | | | | | | | | |
| Schematic | 9 | 169 | −0.34 | 0.36 | 7 | 116 | −0.15 | 0.19 | 8 | 144 | −0.48 | 0.29 | 9 | 163 | 0.47 | 0.26 |
| Rather realistic | 7 | 152 | 0.01 | 0.47 | 9 | 185 | 0.03 | 0.36 | 4 | 79 | −0.33 | 0.21 | 7 | 119 | 0.47 | 0.28 |
| Photo-realistic | 26 | 618 | 0.02 | 0.37 | 27 | 854 | −0.09 | 0.27 | 25 | 827 | −0.36 | 0.29 | 30 | 893 | 0.44 | 0.27 |
| Dimension | | | | | | | | | | | | | | | | |
| Two-dimensional | 38 | 871 | −0.07 | 0.41 | 38 | 1,059 | −0.08 | 0.27 | 34 | 1,000 | −0.37 | 0.29 | 42 | 1,109 | 0.44 | 0.27 |
| Three-dimensional | 5 | 78 | 0.27 | 0.24 | 6 | 106 | −0.17 | 0.34 | 3 | 50 | −0.57 | 0.03 | 4 | 66 | 0.54 | 0.28 |
| Modality | | | | | | | | | | | | | | | | |
| Visual | 21 | 452 | −0.19 | 0.43 | 17 | 334 | 0.02 | 0.31 | 17 | 369 | −0.45 | 0.29 | 21 | 440 | 0.50 | 0.23 |
| Visual plus auditory | 3 | 46 | 0.28 | 0.32 | 4 | 74 | 0.22 | 0.23 | 3 | 58 | −0.31 | 0.32 | 5 | 98 | 0.61 | 0.17 |
| Text annotation | | | | | | | | | | | | | | | | |
| Not annotated | 35 | 789 | −0.02 | 0.39 | 28 | 640 | 0.01 | 0.29 | 29 | 711 | −0.42 | 0.26 | 37 | 863 | 0.48 | 0.27 |
| Annotated | 8 | 160 | −0.13 | 0.52 | 16 | 525 | −0.21 | 0.21 | 8 | 339 | −0.28 | 0.32 | 9 | 312 | 0.35 | 0.26 |
| Task characteristics | | | | | | | | | | | | | | | | |
| Task complexity | | | | | | | | | | | | | | | | |
| Viewing task | 5 | 94 | 0.13 | 0.48 | 7 | 160 | −0.05 | 0.37 | 2 | 44 | −0.43 | 0.38 | – | – | – | – |
| Detection task | 8 | 177 | −0.29 | 0.41 | 11 | 416 | −0.20 | 0.17 | 11 | 416 | −0.30 | 0.34 | 11 | 337 | 0.33 | 0.27 |
| Decision task | 27 | 605 | 0.01 | 0.39 | 20 | 462 | 0.01 | 0.29 | 19 | 453 | −0.47 | 0.20 | 26 | 637 | 0.46 | 0.28 |
| Problem-solving task | 3 | 73 | −0.08 | 0.13 | 6 | 127 | −0.11 | 0.25 | 4 | 99 | −0.31 | 0.29 | 7 | 157 | 0.58 | 0.18 |

**Table 7** (continued)

| | Number of fixations | | | | Fixation duration | | | | Response time | | | | Performance accuracy | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Moderator variable | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD | $k$ | $N$ | $r_c$ | SD |
| Time-on-task | | | | | | | | | | | | | | | | |
| Limited | 32 | 760 | 0.04 | 0.37 | 35 | 1,013 | −0.10 | 0.28 | 24 | 808 | −0.34 | 0.28 | 31 | 893 | 0.40 | 0.28 |
| Unlimited | 11 | 189 | −0.36 | 0.43 | 9 | 152 | −0.04 | 0.27 | 13 | 242 | −0.49 | 0.29 | 15 | 282 | 0.59 | 0.18 |
| Task control | | | | | | | | | | | | | | | | |
| System-controlled | 40 | 917 | −0.04 | 0.42 | 38 | 1,063 | −0.09 | 0.28 | 33 | 976 | −0.37 | 0.29 | 39 | 1,043 | 0.43 | 0.27 |
| User-controlled | 3 | 32 | 0.03 | 0.20 | 6 | 102 | −0.05 | 0.26 | 4 | 74 | −0.41 | 0.23 | 7 | 132 | 0.56 | 0.20 |
| Domain characteristics | | | | | | | | | | | | | | | | |
| Professional domain | | | | | | | | | | | | | | | | |
| Sports | 31 | 680 | 0.00 | 0.41 | 22 | 519 | −0.02 | 0.29 | 21 | 538 | −0.46 | 0.25 | 28 | 704 | 0.48 | 0.27 |
| Team sports | 19 | 463 | −0.04 | 0.44 | 13 | 332 | −0.05 | 0.31 | 13 | 325 | −0.45 | 0.29 | 17 | 423 | 0.49 | 0.31 |
| One-on-one sports | 8 | 174 | 0.09 | 0.31 | 9 | 185 | 0.06 | 0.22 | 7 | 193 | −0.44 | 0.14 | 9 | 254 | 0.46 | 0.21 |
| Solo sports | 4 | 43 | 0.04 | 0.36 | – | – | – | – | – | – | – | – | 2 | 27 | 0.55 | 0.08 |
| Medicine | 4 | 110 | −0.21 | 0.38 | 4 | 75 | −0.06 | 0.14 | 5 | 115 | −0.29 | 0.18 | 6 | 101 | 0.41 | 0.30 |
| Transportation | 4 | 102 | 0.05 | 0.41 | 11 | 455 | −0.19 | 0.22 | 5 | 297 | −0.28 | 0.33 | 6 | 260 | 0.30 | 0.26 |
| Other | 4 | 57 | −0.30 | 0.36 | 6 | 108 | 0.01 | 0.40 | 6 | 100 | −0.31 | 0.33 | 6 | 110 | 0.59 | 0.07 |

Corresponding 99% confidence intervals are available from the corresponding author.

**Table 8** Cumulative evidence concerning the assumptions of three theories of visual expertise

| Theory | Level | Result | Effect size |
|---|---|---|---|
| Theory of long-term working memory | E–N | Experts had shorter fixation durations than novices | −0.09[a] |
| | E–I | Fixation duration of experts and intermediates did not differ | 0.00 |
| | I–N | Intermediates had longer fixation durations than novices | 0.04 |
| Information-reduction hypothesis | E–N | Experts had more fixations of longer duration than novices on task-relevant areas | 0.53[a]<br>0.27[a] |
| | E–N | Experts had fewer fixations of shorter duration than novices on task-redundant areas | −0.31[a]<br>−0.43[a] |
| | E–I | Experts had longer fixation durations on task-relevant and shorter fixation durations on task-redundant areas than intermediates | 0.07[a]<br>−0.01[a] |
| | I–N | Intermediates had longer fixation durations on task-relevant and shorter fixation durations on task-redundant areas than novices | 0.03[a]<br>−0.26[a] |
| Holistic model of image perception | E–N | Experts had shorter times to first fixate task-relevant areas and a longer saccade length than novices | −0.31[a]<br>0.30[a] |
| | E–I | Experts had shorter times to first fixate task-relevant areas and a longer saccade length than intermediates | −0.23[a]<br>0.38[a] |
| | I–N | Novices had shorter times to first fixate task-relevant areas and a longer saccade length than intermediates | −0.01<br>0.12 |

*E–N* experts and novices, *E–I* experts and intermediates, *I–N* intermediates and novices

[a] Assumption of theory verified. Effect sizes are corrected correlation estimates

While the latter finding was based on a small cell size, it can be also related to a nonmonotonicity of development. Specifically, in their study on medical diagnosis, Lesgold *et al*. (1988, p. 334) noted that “performance in diagnosing radiographic films is not always a monotone function of experience.” This statement is not supported by our performance data, which indicated that performance is a monotone function of experience. However, we did find nonmonotonicities on a perceptual level. Particularly, evidence of longer fixation duration for intermediates than for novices suggests an attempt to apply a complex knowledge base, which is not yet fully automatized for intermediates. In general, a total of 92 effect sizes from eye-tracking research confirmed assumptions of the theory of long-term working memory (Ericsson and Kintsch 1995) that experts encode and retrieve information more rapidly than non-experts: Experts’ rapid information processing was reflected in shorter fixation durations.

The primary meta-analysis aimed at testing a second theory as well. The meta-analytic findings tended to confirm assumptions of the information-reduction hypothesis (Haider and Frensch 1999) that expertise optimizes the amount of processed information by a neglect of task-irrelevant information and an active focusing on task-relevant information, which is accomplished through strategic considerations to allocate attentional resources (hypotheses 2a and 2b). A total of 69 effect sizes from eye-tracking research substantiated

Haider and Frensch's (1999) initial theorizing: Selective attention was reflected in a higher number of fixations on task-relevant information.

Finally, the primary meta-analysis aimed at testing a third theory. Results are in line with the holistic model of image perception (Kundel *et al.* 2007; hypotheses 3a and 3b), suggesting that experts are able to extract information from widely distanced and parafoveal regions (Reingold *et al.* 2001). In general, 29 effect sizes from eye-tracking research confirmed the assumption that expertise extends the visual span: Experts' parafoveal processing was reflected in longer saccades and shorter times to first fixate on areas of task relevance.

In addition to the eye movement parameter, our meta-analysis complemented these findings with an examination of reaction time and performance accuracy. A total of 50 effect sizes confirmed the superior speed of expert task processing, and 57 effect sizes indicated that experts were more accurate in task performance than non-experts. It seems safe to conclude that the systematic eye movement differences are related to experts' reproducibility of domain-specific task superiority.

The average number of participants in expertise research using eye-tracking methodology—11 experts, 10 intermediates, and 12 novices—indicates the value of meta-analytic cumulation to correct individual study findings for the bias of sampling error. Moreover, analysis of demographic variables in Table 5 provided additional evidence that the 10-year rule is a weak indicator for expertise. For example, the mean number of years of experience was higher for transportation intermediates than for transportation experts, so it seems that experience is a necessary but insufficient indicator for expert performance (Billett 2009; Ericsson and Lehmann 1996; Feltovich *et al.* 2006).

### Discussion of meta-analytic moderator analyses

The results of the meta-analytic moderator analyses illustrated boundary conditions of expertise differences in the comprehension of visualizations. In particular, the size of the difference was moderated by visualization characteristics. First, because different aspects of the visualizations used in primary research reduced extraneous processing demands in working memory and fostered generative processing of essential material, expertise differences varied. Consistent with our hypotheses, smaller performance differences were found when the visualization was static rather than dynamic (hypothesis 4a), two-dimensional rather than three-dimensional (hypothesis 4c), and annotated by text rather than without annotation (hypothesis 4e). Contrary to expectations, larger performance differences were found for schematic rather than realistic visualizations (hypothesis 4b) that used visual plus auditory modality rather than visual modality only (hypothesis 4d). There are at least two possible explanations for the latter findings. Novices may have been unfamiliar with schematic representations of realistic scenes; the unknown level of abstraction may have induced high cognitive load in working memory. Another explanation may be that the findings resulted from second-order sampling error, for some of the categories on realism and modality are among those with the smallest cell sizes. Overall, Table 7 illustrates that visualization characteristics moderated the size of expertise differences in eye movements and performance variables.

The meta-analytic moderator analyses indicated that expertise differences were also influenced by task characteristics. Consistent with our hypotheses, smaller differences were found for less complex rather than more complex tasks (hypothesis 5a) with limited rather than unlimited time-on-task (hypothesis 5b). Contrary to expectations, the differences were larger for tasks controlled by the user (hypothesis 5c). Although we assumed that user-paced rather than system-paced visualizations allow novices to regulate visuospatial processing demands in working memory (Mayer 2009; Spanjers *et al*. 2010), it seems that complex visual environments may be problematic, causing disorientation and extraneous processing overload. An alternative explanation to hypothesis 5c is that, when dealing with transient information (spoken text in dual-modality presentations or animated visualizations), the length of the instructional episode may significantly influence working memory load and potentially override the expected effect of learner control and other factors.

Finally and in accordance with our hypothesis, differences varied as a function of the professional domain (hypothesis 6). A clear pattern emerged for response time, with generally stronger differences for sport domains. This difference is likely the result of task affordances in sport environments. For example, rapid action generation within seconds is more often found in soccer than in the viewing of art pictures (Vickers 2007). Within sport domains, particularly in timing and tactical tasks, specific gaze behavior is a characteristic of higher levels of sport performance (cf. theory of the quiet eye, Vickers 2007), so visual search behavior tends to be moderated by task affordances across and within professional domains.

## Implications for theory development

The findings of both the primary meta-analysis and the meta-analytic moderator analysis bear on the development of theories on visual expertise in graphics comprehension. It is evident from the meta-analytic eye-tracking data reported in the present study that rapid information processing, selective attention allocation, and extension of the visual span constitute important components for any theory of visual expertise. In addition, the limited-capacity assumption that characterizes theories on multimedia learning (Mayer 2009; Sweller 1994) needs to be reconsidered when accounting for experts' superior processing resources of domain-specific material. Boundary conditions of the limited-capacity assumption in cases of expert learners are reflected in the prior knowledge principle (Mayer 2009) and the expertise reversal effect (Sweller *et al*. 2011). The present study outlines potential intersections between the three theories reviewed (Ericsson and Kintsch 1995; Haider and Frensch 1999; Kundel *et al*. 2007) in the comprehension of visualizations. Further specification and research efforts are needed with respect to intermediates. While expert–novice differences seem well examined, there has been scarcity of research on intermediate participants. Two conclusions can be given with regard to our data. First, intermediates have been found to have higher accuracy and shorter reaction times than novices and smaller accuracy and longer reaction times than experts. This finding is hardly surprising. Maybe more surprising is the second conclusion, which suggests shorter saccade length and longer times to first fixate relevant areas were found for intermediates in comparison to novices. Although this finding may be influenced by the small cell size, it also indicates that intermediates have not yet acquired the ability of parafoveal processing (Kundel *et al*. 2007; Reingold *et al*. 2001). It follows that rapid processing owing to retrieval cues and learned selectivity may develop first, and at later stages of expertise acquisition, the development of an extended visual span may follow. We shall note, however, that contrastive between-group differences are at best a limited indicator for longitudinal human development. Because of extant evidence on cross-sectional differences and a scarcity on developmental trajectories of intermediates over time, future research may be directed to further longitudinal examination of how expertise changes intermediates' cognitive architecture over extended time frames. For such investigations, eye movement recordings may provide useful insights about how expertise influences stimulus encoding (Hyönä 2010; Just and Carpenter 1984). Notwithstanding the utility of the *eye–mind hypothesis*, a cautionary note seems appropriate: eye-tracking methodology has limitations in producing indicators of expertise differences. As Hyönä (2010, p. 173) noted, "gaze behavior can serve as an index of current attentional processes only as long as the available visual environment in front of our eyes is pertinent to the task we would like to study."

## Implications for the design of learning environments

The analysis of eye movement differences has some implications for the design of learning environments. The effect sizes shown in Table 8 demonstrate training needs for novices, particularly with regard to directing attention to areas with high information value. One approach to meet the identified training needs is to design technological learning environments. Specifically, a replay of the eye movements of experts, superimposed on the screen showing the visualization, can be used to model the eye movements of novices. Following another person's gaze to model attentional resource allocation is a mechanism well documented with samples ranging from infants (Meltzoff *et al*. 2010) to college students (Nalanagula *et al*. 2006). In professional domains that heavily depend on visual information, such as in medical image diagnosis (Kundel *et al*. 2007) or aviation security (Liu *et al*. 2007), the gaze following mechanism can be used to model the eye movements of novices, which would be beneficial for several reasons. First, trainees may learn *what* to focus on. Results of the primary meta-analysis demonstrated that experts fixate on areas with high information value, so replaying the gaze behavior of experts may help novices detect task-relevant areas and perceptually ignore redundant areas (Nalanagula *et al*. 2006). Second, trainees may learn *in which order* to focus. Analysis of three-dimensional visualizations indicated that experts had more fixations than novices, so modeling the patterns of visual search may provide perceptual procedure cues on how to navigate in complex environments (Wilson *et al*. 2010). Third, attentional guidance may improve not only detection and the order of visual search, but also reasoning (Henderson *et al*. 2010). Grant and Spivey (2003) indicated that cognitive processing is sometimes the result of attention and eye movements, so expert gaze replay may improve skilled thinking (Kuhn 2009). Finally, interacting with digital media, such as computer-based gaze replays, can increase trainee motivation, interest, and engagement, particularly through situational affordances, which function as important precursors for work-related learning (Hidi 2006; Gnaur 2010). To summarize, analysis of eye movement differences may inform the design of learning environments to include viewing the scan paths of experts for directing the attentional resources of novices. Further efforts are needed to investigate how the effect sizes of the present meta-analysis can be used to support expert development, particularly in those professions that heavily depend on visual information (Kundel *et al*. 2007; Liu *et al*. 2007).

## Study limitations

This study has some limitations that should be noted. One limitation is that this meta-analysis connected original studies from different fields of expertise and the variation of the nature of tasks used in the studies is substantial. Even though this kind of aggregate analysis is relevant for analyzing general hypotheses emerging from expertise theories, it does not allow as rigor testing of specific hypothesis as meta-analyses based on a coherent set of studies. This should be taken into account in interpreting the results.

Another limitation is that the correlation estimates in Tables 6 and 7 were corrected for sampling error. This decision was based on the frequent reporting and availability of sample size information. However, the original research reports may be affected by additional biases, such as extraneous factors introduced by study procedure (Hunter and Schmidt 2004). Although the identification and estimation of moderators sought to reduce this bias, the corrected correlation estimates may be somewhat greater than those reported here.

An additional limitation is that some of the expert–novice differences in the primary meta-analysis were based on small sample sizes. A related concern is the small cell size for a few of the visualization, task, and domain characteristics in the meta-analytic moderator analysis. However, some authors have noted that correcting for bias at a small scale mitigates sampling error compared to uncorrected estimates in individual studies (Hunter and Schmidt 2004; Rosenthal and DiMatteo 2001). Still, although most of the cells contained sample sizes in the hundreds, some did contain fewer, which indicates underestimation of sampling error in those few cases.

A further potential limitation is the fact that the study addressed three groups of moderator effects. It was implicitly assumed that visualization, task, and domain characteristics are among the three most dominant sources for effect size heterogeneity in graphics comprehension. Nonetheless, the total number of potential moderator variables likely exceeds three. Results of the meta-analytic moderator analysis are therefore limited in their generalizability across the full range of possible study conditions.

Moreover, the primary literature used different criteria to categorize participants as expert or novice. These criteria include performance efficiency (Van Gog *et al.* 2005), social recognition (Vickers 1988), group membership (Ripoll *et al.* 1995), and years of experience (Williams and Davids 1997); a complete description of criteria used in the primary literature is offered in the "Appendix." Of course, it is reasonable to assume that the use of different indicators used by researchers to select samples of experts, intermediates, and novices can contribute to the variability of expertise differences. While an inquiry into expertise criterion as a moderator was beyond the scope of the present study, researchers are invited to use our review of the literature (see the "Appendix") for addressing this question. Therefore, future research may want to estimate the extent to which researchers' selection of expertise criterion may moderate expertise-related between-group differences in the comprehension of visualizations. However, it is not the absolute level of expertise that is important in this study but the relative expertise demonstrated as the difference between novices and experts (Chi 2006).

Finally, the study covered eight eye movement parameters. Although an analysis of the difference of these eight variables under different moderator conditions clearly goes beyond previous meta-analytic attempts, selection of the eye-tracking measures was eclectic. More parameters exist that would warrant inclusion in the meta-analysis. However, this limitation can be addressed only by additional original research reports that consider different eye movement parameter in expertise research. Therefore and because of the limitations discussed previously, this meta-analysis represents only a first step toward improving our understanding of expertise differences in the comprehension of visualizations.

## Conclusion

As noted at the outset, expertise in the comprehension of visualizations has gained growing attention over the last few years. Eye-tracking research on expertise differences were not always in agreement with theoretical arguments. This study sought to evaluate meta-analytically three theories on expertise in visual domains by focusing on eight eye movement and two performance variables; by cumulating 73 independent data sources with 296 effect sizes from 819 experts, 187 intermediates, and 893 novices; and by examining nine theory-driven moderator variables on the size of the expertise differences. The findings inform the development of theories of visual expertise to include expert capabilities concerning rapid information processing, selective attention allocation, and extension of the visual span. Future research is encouraged to extend the first steps reported here to the examination of how expert gaze replay can be implemented in learning environments to support expert development in vision-intensive professions.

**Acknowledgments** This research was supported in part by grants from the Academy of Finland and from the Doctoral Program for Multidisciplinary Research on Learning Environments (OPMON).

## Appendix

**Table 9** Eye-tracking studies of expertise differences in the comprehension of visualizations

| Author(s) (year) | Domain | Visualization | Task | Level | Expertise criterion |
|---|---|---|---|---|---|
| Abernethy and Russel (1987) | Badminton | Video film | Anticipation of the landing position of opponent's stroke | E–N | Previous participation in VIIth Commonwealth Games |
| Abernethy (1990) (experiment 1) | Squash | Video film | Anticipation of the landing position of opponent's stroke | E–N | Group membership; ranking position |
| Amadieu et al. (2009) | Virology | Digital concept maps | Orienting in concept maps | E–N | Prior knowledge |
| Augustyniak and Tadeusiewicz (2006) | Cardiology | ECG | ECG interpretation | E–N | Professional specialization; years of experience |
| Bard and Fleury (1976) | Basketball | Schematic slides | Deciding the best move | E–N | Years of active competition in basketball |
| Bard et al. (1980) | Gymnastics | Video film | Error detection in routines on the balance beam | E–N | Years of experience as gymnastics judges; national certification |
| Bednarik and Tukiainen (2007) | Java programming | Software debugging environment | Debugging | E–N | Professional background; experience with Java programming |
| Bellenkes et al. (1997) | Aviation | Flight simulator | Performing flight maneuvers | E–N | Professional background; number of flight hours |
| Bertrand and Thullier (2009) | Soccer | Sequence of static frames from a film | Anticipation of dribble direction | E–N | Years of experience in playing soccer |
| Cauraugh et al. (1993) | Tennis | Video film | Decision about the intentions of a filmed opponent | E–N | Former NCAA champions |
| Chapman and Underwood (1998) | Car driving | Video film | Detecting hazardous events | E–N | Years of having a driving license; miles driven |
| Chapman and Underwood (1999) | Car driving | Video film | Detecting hazardous events | E–N | Years of having a driving license; miles driven; number of accidents |
| Charness et al. (2001) | Chess | Schematic pictures of chess positions | Deciding the best move | E–I | Ratings of Chess Federation of Canada |

**Table 9** (continued)

| Author(s) (year) | Domain | Visualization | Task | Level | Expertise criterion |
|---|---|---|---|---|---|
| Cooper et al. (2009) | Radiology | Multidimensional brain scans | Stroke detection | E–N | Experience; group membership |
| Cooper et al. (2010) | Radiology | Multidimensional brain scans | Stroke detection | E–I–N | Experience; group membership |
| Crundall et al. (1999) | Car driving | Video film | Detecting hazardous events | E–I–N | Years since having a driving license |
| Crundall et al. (2003) | Car driving | Video film | Detecting hazardous events | E–I–N | Years since having a driving license; annual mileage; professional background |
| Crundall et al. (2005) | Car driving | Video film | Detecting hazardous events | E–I–N | Years since having a driving license; annual mileage; professional background |
| de Groot and Gobet (1996) (Chapter 6) | Chess | Photographs | Reproducing board positions | E–N | International master or grandmaster |
| Dyer et al. (2006) | Forensics | Photographs | Deciding whether a signature was genuine or forged | E–N | Years of professional experience; qualification to present expert evidence regarding signatures |
| Goulet et al. (1989) (experiment 1) | Tennis | Video film | Identification of the type of serve | E–N | (Previously) ranked among the top 40 in Québec |
| Helsen and Pauwels (1990) | Soccer | Static slides | Deciding the next action | E–N | Years of active competition |
| Helsen and Pauwels (1992) | Soccer | Video film | Deciding the next action | E–N | Years of active competition |
| Helsen and Starkes (1999) (experiment 2) | Soccer | Static slides | Deciding the next action | E–I | Years of active competition |
| Helsen and Starkes (1999) (experiment 3) | Soccer | Video film | Deciding the next action | E–I | Years of active competition |
| Hermans and Laarni (2003) | Cartography | Schematic screen maps | Target detection | E–N | Professional background |
| Huestegge et al. (2010) | Car driving | Photographs | Detecting hazardous events | E–N | Years of driving experience; miles driven |
| Jarodzka et al. (2010) | Biology | Video film | Classification of fish locomotion | E–N | Years of practical experience; interest |
| Kasarskis et al. (2001) | Aviation | Flight simulator | Landing an aircraft | E–N | Professional background; logged flight hours |
| Kato and Fukuda (2002) | Baseball | Video film | Viewing pitches | E–N | Group membership |
| Konstantopoulos (2009) (experiment 4) | Car driving | Driving simulator | Driving a car | E–N | Years of driving experience; professional background |
| Konstantopoulos (2009) (experiment 5) | Car driving | Driving simulator | Driving a car | E–N | Years of driving experience; professional background |
| Kristjanson and Antes (1989) | Arts | Paintings | Viewing paintings | E–N | Group membership; engagement with art |
| Krupinski (2005) | Radiology | X-ray mammogram | Lesion detection | E–N | Reading volume |

**Table 9** (continued)

| Author(s) (year) | Domain | Visualization | Task | Level | Expertise criterion |
|---|---|---|---|---|---|
| Krupinski et al. (2006) | Pathology | Static breast biopsy slides | Selecting the top 3 locations to zoom on | E–I–N | Years of pathology practice experience; board certification |
| Kundel et al. (2007) | Radiology | X-ray mammogram | Lesion detection | E–I–N | Post hoc performance |
| Laurent et al. (2006) (experiment 1) | Basketball | Schematic pictures of basketball configurations | Judgment whether stimuli were the same or different | E–N | Years of deliberate practice; competition at national level |
| Laurent et al. (2006) (experiment 2) | Basketball | Schematic pictures of basketball configurations | Judgment whether stimuli were the same or different | E–N | Years of deliberate practice; competition at national level |
| Litchfield et al. (2008) | Radiology | Chest X-ray | Nodule detection | E–N | Years of experience; group membership |
| Liu et al. (2007) | Airport security | X-ray luggage scans | Detecting thread items | E–N | Group membership |
| Manning et al. (2006) | Radiology | Chest X-ray | Nodule detection | E–I–N | Experience; group membership |
| McRobert et al. (2009) | Cricket | Video film | Prediction of the flight path of opponent's delivery | E–N | Years of playing experience; number of competitive matches; professional background |
| Moran et al. (2002) (experiment 2) | Equestrian riding | Fence photographs and schematic maps | Viewing fences in a "walk the course" | E–I–N | Years of riding experience; participation at Olympic and World Championships |
| Moreno et al. (2002) | Gymnastics | Video film | Error detection in gymnastic routines | E–N | Professional background |
| Moreno et al. (2006) | Swimming | Video film | Error detection in swim moves | E–N | Experience in underwater viewing; coaching experience |
| Nodine et al. (1996) | Radiology | X-ray mammogram | Lesion detection | E–I–N | Levels of training; reading volume |
| Nodine et al. (2002) | Radiology | X-ray mammogram | Lesion detection | E–N | Levels of training; reading volume |
| North et al. (2009) | Soccer | Video film | Anticipation of the ball outcome destination | E–N | Years of playing experience; weekly training hours; professional background |
| North et al. (2009) | Soccer | Schematic point-light film | Anticipation of the ball outcome destination | E–N | Years of playing experience; weekly training hours; professional background |
| Raab and Johnson (2007) | Handball | Video film | Generating options of the player | E–N | National champions; training amount and content |
| Reingold et al. (2001) (experiment 1) | Chess | Schematic pictures of chess positions | Detecting a changing piece | E–I–N | Ratings of Chess Federation of Canada |
| Reingold et al. (2001) (experiment 2) | Chess | Schematic pictures of chess positions | Check detection | E–I–N | Ratings of Chess Federation of Canada |
| Ripoll et al. (1993) | Boxing | Video film | Anticipation of different boxing situations | E–I–N | Professional background |
| Ripoll et al.(1995) (experiment 2) | Boxing | Video film | Anticipation of different boxing situations | E–I–N | Members of national team |

**Table 9** (continued)

| Author(s) (year) | Domain | Visualization | Task | Level | Expertise criterion |
|---|---|---|---|---|---|
| Savelsbergh et al. (2002) | Soccer | Video film | Anticipating direction of penalty kicks | E–N | Professional background; years of active competition |
| Savelsbergh et al. (2005) | Soccer | Video film | Anticipating direction of penalty kicks | E–N | Post hoc performance |
| Schriver et al. (2008) | Aviation | Flight simulator | Troubleshooting during flight | E–N | Flight hours; certification; general test of aviation knowledge |
| Singer et al. (1996) | Tennis | Video film | Decision about the intentions of a filmed opponent | E–N | National champions; membership in a top 10 collegiate team |
| Underwood et al. (2002) | Car driving | Video film | Driving a car | E–N | Years of driving experience; miles driven |
| Vaeyens et al. (2007) | Soccer | Video film | Deciding the next action | E–N | Performance ranking |
| Van Gog et al. (2005) | Physics | Schematic picture of electrical circuits | Troubleshooting | E–N | Performance efficiency |
| Vickers (1988) (study 1) | Gymnastics | Photographs of gymnastic sequences | Viewing the photographs | E–I–N | National ranking; supervisory nomination; years in gymnastics competition |
| Vogt and Magnussen (2007) | Arts | Art pictures | Viewing art pictures | E–N | Group membership; years of training |
| Ward et al. (2002) | Tennis | Video film | Anticipation of the ball outcome destination following groundstrokes | E–N | Years of playing experience; number of competitive matches; regular training; professional background |
| Ward et al. (2002) | Tennis | Schematic point-light film | Anticipation of the ball outcome destination following groundstrokes | E–N | Years of playing experience; number of competitive matches; regular training; professional background |
| Williams et al. (1994) | Soccer | Video film | Anticipation of pass destination | E–N | Years of playing experience; number of competitive matches; professional background |
| Williams and Davids (1997) (experiment 1) | Soccer | Video film | Anticipation of pass destination | E–N | Years of playing experience; number of competitive matches; professional background |
| Williams and Davids (1997) (experiment 2) | Soccer | Video film | Anticipation of pass destination | E–N | Years of playing experience; number of competitive matches; professional background |
| Williams and Davids (1998) (experiment 1A) | Soccer | Video film | Anticipation of pass destination | E–N | Years of playing experience; number of competitive matches; professional background |
| Williams and Davids (1998) (experiment 1B) | Soccer | Video film | Anticipation of dribble direction | E–N | Years of playing experience; number of competitive matches; professional background |
| Williams and Elliott (1999) | Karate | Video film | Anticipation of attacks | E–N | Minimum of 3 years of training; regular sparring practice; competition experience |

**Table 9** (continued)

| Author(s) (year) | Domain | Visualization | Task | Level | Expertise criterion |
|---|---|---|---|---|---|
| Williams et al. (2002) (experiment 1) | Tennis | Video film | Anticipation of the ball outcome destination following groundstrokes | E–N | Years of playing experience; number of competitive matches; professional background |
| Wilson et al. (2010) | Laparoscopy | Surgical simulator | Eye–hand coordination | E–N | Number of laparoscopic procedures led |

*E–N* experts and novices, *E–I* experts and intermediates, *E–I–N* experts, intermediates, and novices, *ECG* electrocardiograms

## References

*Abernethy, B. (1990). Expertise, visual search, and information pick-up in squash. *Perception, 19*, 63–77.

*Abernethy, B., & Russel, D. G. (1987). The relationship between expertise and visual search strategy in a racquet sport. *Human Movement Science, 6*, 283–319.

*Amadieu, F., Van Gog, T., Paas, F., Tricot, A., & Mariné, C. (2009). Effects of prior knowledge and concept-map structure on disorientation, cognitive load, and learning. *Learning and Instruction, 19*, 376–386.

*Augustyniak, P., & Tadeusiewicz, R. (2006). Assessment of electrocardiogram visual interpretation strategy based on scanpath analysis. *Physiological Measurement, 27*, 597–608.

*Bard, C., & Fleury, M. (1976). Analysis of visual search activity during sport problem situations. *Journal of Human Movement Studies, 3*, 214–222.

*Bard, C., Fleury, M., Carrière, L., & Hallé, M. (1980). Analysis of gymnastics judges' visual search. *Research Quarterly for Exercise and Sport, 51*, 267–273.

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & D. E. Mark (Eds.), *The cognitive neuroscience of working memory* (pp. 57–80). Oxford: Oxford University Press.

Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in meta-analysis of correlations. *Journal of Applied Psychology, 87*, 583–589.

*Bednarik, R., & Tukiainen, M. (2007). Validating the restricted focus viewer: A study using eye-movement tracking. *Behavior Research Methods, 39*, 274–282.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space and Environmental Medicine, 48*, 569–579.

*Bertrand, C., & Thullier, F. (2009). Effects of player position task complexity in visual exploration behavior in soccer. *International Journal of Sport Psychology, 40*, 306–323.

Billett, S. (2009). Personal epistemologies, work, and learning. *Educational Research Review, 4*, 210–219.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*, 182–197.

Campbell, D. J. (1986). Task complexity: A review and analysis. *Academy of Management Review, 13*, 40–52.

*Cauraugh, J. H., Singer, R. N., & Chen, D. (1993). Visual scanning and anticipation of expert and beginning tennis players. In S. Serpa (Ed.), *Proceedings of the World Congress of Sport Psychology* (pp. 336–340). Lisbon: International Society of Sport Psychology.

*Chapman, P. R., & Underwood, G. (1998). Visual search of driving situations: Danger and experience. *Perception, 27*, 951–964.

*Chapman, P. R., & Underwood, G. (1999). Visual search of dynamic scenes: Event types and the role of experience in viewing driving situations. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 369–393). Amsterdam: Elsevier.

*Charness, N., Reingold, E. M., Pomplun, M., & Stampe, D. M. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition, 29*, 1146–1152.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21–30). Cambridge: Cambridge University Press.

*Cooper, L., Gale, A., Darker, I., Toms, A., & Saada, J. (2009). Radiology image perception and observer performance: How does expertise and clinical information alter interpretation? Stroke detection explored through eye-tracking. In B. Sahiner & D. J. Manning (Eds.), *Medical Imaging 2009: Image perception, observer performance, and technology assessment Vol. 7263* (pp. 72630K–72630K-12). Bellingham: SPIE.

*Cooper, L., Gale, A., Saada, J., Gedela, S., Scott, H., & Toms, A. (2010). The assessment of stroke multidimensional CT and MR imaging using eye movement analysis: Does modality preference enhance observer performance? In D. J. Manning & C. K. Abbey (Eds.), *Medical Imaging 2010: Image perception, observer performance, and technology assessment Vol. 7627* (pp. 76270B–76270B-12). Bellingham: SPIE.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

*Crundall, D., Underwood, G., & Chapman, P. (1999). Driving experience and the functional field of view. *Perception, 28*, 1075–1087.

*Crundall, D., Chapman, P., Phelps, N., & Underwood, G. (2003). Eye movements and hazard perception in police pursuit and emergency response driving. *Journal of Experimental Psychology: Applied, 9*, 163–174.

*Crundall, D., Chapman, P., France, E., Underwood, G., & Phelps, N. (2005). What attracts attention during police pursuit driving? *Applied Cognitive Psychology, 19*, 409–420.

*de Groot, A. D., & Gobet, F. (1996). *Perception and memory in chess. Heuristics of the professional eye*. Assen: Van Gorcum.

Dwyer, F. M., & Joseph, J. (1984). The effects of prior knowledge, presentation mode, and visual realism on student achievement. *Journal of Experimental Education, 52*, 110–121.

*Dyer, A. G., Found, B., & Rogers, D. (2006). Visual attention and expertise for forensic signature analysis. *Journal of Forensic Sciences, 51*, 1397–1404.

Eilam, B., & Poyas, Y. (2008). Learning with multiple representations: Extending multimedia learning beyond the lab. *Learning and Instruction, 18*, 368–378.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211–245.

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptations to task constraints. *Annual Review of Psychology, 47*, 273–305.

Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.

Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 41–67). Cambridge: Cambridge University Press.

Gibson, J. J. (1986). *The ecological approach to visual perception*. New York: Psychology.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction, 15*, 313–331.

Gnaur, D. (2010). Seizing workplace learning affordances in high-pressure work environments. *Vocations and Learning, 3*, 223–238.

Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*, 585–612.

*Goulet, C., Bard, C., & Fleury, M. (1989). Expertise differences in preparing to return a tennis serve: A visual information processing approach. *Journal of Sport & Exercise Psychology, 11*, 382–398.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science, 14*, 462–466.

Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 25*, 172–190.

Hall, S. M., & Brannick, M. T. (2002). Comparison of two random effects methods of meta-analysis. *Journal of Applied Psychology, 87*, 377–389.

*Helsen, W., & Pauwels, J. M. (1990). Analysis of visual search activity in solving tactical game problems. In D. Brogan (Ed.), *Visual search* (pp. 177–184). London: Taylor & Francis.

*Helsen, W., & Pauwels, J. M. (1992). A cognitive approach to visual search in sport. In D. Brogan (Ed.), *Visual search 2* (pp. 379–388). London: Taylor & Francis.

*Helsen, W., & Starkes, J. L. (1999). A multidimensional approach to skilled perception and performance in sport. *Applied Cognitive Psychology, 13*, 1–27.

Henderson, A. J., Alexander, H., Haywood, A., Stapleton, P., Cooke, M., Patterson, E., et al. (2010). Promoting and providing expert guidance in work-intensive clinical settings. *Vocations and Learning, 3*, 141–156.

*Hermans, O. & Laarni, J. (2003). Searching information from screen maps. *Paper presented at the 9th Scandinavian Research Conference on Geographical Information Science*, Espoo.

Hidi, S. (2006). Interest: A unique motivational variable. *Educational Research Review, 1*, 69–82.

*Huestegge, L., Skottke, E.-M., Anders, S., Müsseler, J., & Debus, G. (2010). The development of hazard perception: Dissociation of visual orientation and hazard processing. *Transportation Research Part F: Traffic Psychology and Behaviour, 13*, 1–8.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed.). Thousand Oaks: Sage.

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176.

*Jarodzka, H., Scheiter, K., Gerjets, P., & Van Gog, T. (2010). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*, 146–154.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 151–182). Hillsdale: Erlbaum.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior, 24*, 852–861.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*, 1–11.

*Kasarskis, P., Stehwien, J., Hickox, J., & Aretz, A. (2001). *Comparison of expert and novice scan behaviors during FR flight*. Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH.

*Kato, T., & Fukuda, T. (2002). Visual search strategies of baseball batters: Eye movements during the preparatory phase of batting. *Perceptual and Motor Skills, 94*, 380–386.

*Konstantopoulos, P. (2009). *Investigating drivers' visual search strategies: Towards an efficient training intervention*. Unpublished doctoral thesis, University of Nottingham, Nottingham, England.

*Kristjanson, A. F., & Antes, J. R. (1989). Eye movement analysis of artists and nonartists viewing paintings. *Visual Arts Research, 15*, 21–30.

*Krupinski, E. A. (2005). Visual search of mammographic images: Influence of lesion subtlety. *Academic Radiology, 12*, 965–969.

Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics, 72*, 1205–1217.

*Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., et al. (2006). Eye-movement study and human performance using telepathology virtual slides. *Implications for medical education and differences with experience. Human Pathology, 37*, 1543–1556.

Kuhn, D. (2009). Do students need to be taught how to reason? *Educational Research Review, 4*, 1–6.

*Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology, 242*, 396–402.

*Laurent, E., Ward, P., Williams, A. M., & Ripoll, H. (2006). Expertise in basketball modifies perceptual discrimination abilities, underlying cognitive processes, and visual behaviours. *Visual Cognition, 13*, 247–271.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex domain: Diagnosing X-ray pictures. In M. T. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The nature of expertise* (pp. 311–342). Hillsdale: Erlbaum.

*Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2008). Learning from others: Effects of viewing another person's eye movements while searching for chest nodules. In B. Sahiner & D. J. Manning (Eds.), *Medical Imaging 2008: Image perception, observer performance, and technology assessment Vol. 6917* (pp. 691715–691715-9). Bellingham: SPIE.

*Liu X., Gale A. G., & Song, T. (2007). *Detection of terrorist threats in air passenger luggage: Expertise development*. Paper presented at the 41st Annual IEEE International Carnahan Conference, Ottawa.

Mann, D. T. Y., Williams, A. M., Ward, P., & Janelle, C. M. (2007). Perceptual-cognitive expertise in sport: A meta-analysis. *Journal of Sport & Exercise Psychology, 29*, 457–478.

*Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography, 12*, 134–142.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715–726.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*, 43–52.

*McRobert, A. P., Williams, A. M., Ward, P., & Eccles, D. W. (2009). Tracing the process of expertise in a simulated anticipation task. *Ergonomics, 52*, 474–483.

Meltzoff, A. N., Brooks, R., Shon, A. P., & Rao, R. P. N. (2010). “Social” robots are psychological agents for infants: A test of gaze following. *Neural Networks, 23*, 966–972.

Miller, G. (1956). The magical number seven, plus or minus two: Some limits on our capacity for information processing. *Psychological Review, 63*, 81–97.

*Moran, A., Byrne, A., & McGlade, N. (2002). The effects of anxiety and strategic planning on visual search behaviour. *Journal of Sports Sciences, 20*, 225–236.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*, 358–368.

*Moreno, F. J., Reina, R., Luis, V., & Sabido, R. (2002). Visual search strategies in experienced and inexperienced gymnastic coaches. *Perceptual and Motor Skills, 95*, 901–902.

*Moreno, F. J., Saavedra, J. M., Sabido, R., Luis, V., & Reina, R. (2006). Visual search strategies of experienced and nonexperienced swimming coaches. *Perceptual and Motor Skills, 103*, 861–872.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*, 289–300.

National Research Council (1992). *Combining information: Statistical issues and opportunities for research*. Washington, DC: National Academy of Sciences.

*Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Academic Radiology, 3*, 1000–1006.

*Nodine, C. F., Mello-Thoms, C., Kundel, H. R., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology, 179*, 917–923.

*North, J. S., Williams, A. M., Hodges, N., Ward, P., & Ericsson, K. A. (2009). Perceiving patterns in dynamic action sequences: Investigating the processes underpinning stimulus recognition and anticipation skill. *Applied Cognitive Psychology, 23*, 878–894.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243.

Powers, W. T. (2005). *Behavior: The control of perception*. New Canaan: Benchmark.

*Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied, 13*, 158–170.

*Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players. *Psychological Science, 12*, 48–55.

*Ripoll, H., Kerlirzin, Y., Stein, J.-F., & Reine, B. (1993). Decision making and visual strategies of boxers in a simulated problem-solving situation. In G. d’Ydewalle & J. Van Rensbergen (Eds.), *Perception and cognition: Advances in eye-movement research* (pp. 141–147). Amsterdam: North-Holland.

*Ripoll, H., Kerlirzin, Y., Stein, J.-F., & Reine, B. (1995). Analysis of information processing, decision making, and visual strategies in complex problem solving sport situations. *Human Movement Science, 14*, 325–349.

Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82.

*Savelsbergh, G. J. P., Williams, A. M., Van der Kamp, J., & Ward, P. (2002). Visual search, anticipation and expertise in soccer goalkeepers. *Journal of Sports Sciences, 20*, 279–287.

*Savelsbergh, G. J. P., Van der Kamp, J., Williams, A. M., & Ward, P. (2005). Anticipation and visual search behaviour in expert soccer goalkeepers. *Ergonomics, 48*, 1686–1697.

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494.

Schmidt, F. L., Oh, I.-S., & Hayes, T. L. (2009). Fixed- versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*, 97–128.

*Schriver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors, 50*, 864–878.

*Singer, R. N., Cauraugh, J. H., Chen, D., Steinberg, G. M., & Frehlich, S. G. (1996). Visual search, anticipation, and reactive comparisons between highly-skilled and beginning tennis players. *Journal of Applied Sport Psychology, 8*, 9–26.

Spanjers, I. A. E., Van Gog, T., & Van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students’ learning. *Educational Psychology Review, 22*, 411–423.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Berlin: Springer.

*Underwood, G., Crundall, D., & Chapman, P. (2002). Selective searching while driving: The role of experience in hazard detection and general surveillance. *Ergonomics, 45*, 1–12.

*Vaeyens, R., Lenoir, M., Williams, A. M., & Philippaerts, R. M. (2007). Mechanisms underpinning successful decision making in skilled youth players: An analysis of visual search behaviors. *Journal of Motor Behavior, 39*, 395–408.

*Van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology, 19*, 205–221.

*Vickers, J. N. (1988). Knowledge structures of expert-novice gymnasts. *Human Movement Science, 7*, 47–72.

Vickers, J. N. (2007). *Perception, cognition, and decision training. The quiet eye in action*. Champaign: Human Kinetics.

*Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception, 36*, 91–100.

*Ward, P., Williams, A. M., & Bennett, S. J. (2002). Visual search and biological motion perception in tennis. *Research Quarterly for Exercise and Sport, 73*, 107–112.

*Williams, A. M., & Davids, K. (1997). Assessing cue usage in performance contexts: A comparison between eye-movement and concurrent verbal report methods. *Behavior Research Methods, Instruments, & Computers, 29*, 364–375.

*Williams, A. M., & Davids, K. (1998). Visual search strategy, selective attention, and expertise in soccer. *Research Quarterly for Exercise and Sport, 69*, 111–128.

*Williams, A. M., & Elliott, D. (1999). Anxiety, expertise, and visual search strategy in karate. *Journal of Sport & Exercise Psychology, 21*, 362–375.

*Williams, A. M., Davids, K., Burwitz, L., & Williams, J. G. (1994). Visual search strategies in experienced and inexperienced soccer players. *Research Quarterly for Exercise and Sport, 65*, 127–135.

*Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied, 8*, 259–270.

*Wilson, M., McGrath, J., Vine, S., Brewer, J., Defriend, D., & Masters, R. (2010). Psychomotor control in a virtual laparoscopic surgery training environment: Gaze control parameters differentiate novices from experts. *Surgical Endoscopy, 24*, 2458–2464.

Wood, R. E. (1986). Task complexity: Definition of the construct. *Organizational Behavior and Human Decision Processes, 37*, 60–82.

References marked with an asterisk indicate studies included in the meta-analysis.