Assessment of Visual Literacy –
Contributions of Eye Tracking

Kumulative Inaugural-Dissertation zur Erlangung der Doktorwürde der Fakultät
für Humanwissenschaften der Universität Regensburg

vorgelegt von

Miles Tallon

aus Köln

-2022-

Regensburg 2022

Universität Regensburg

Gutachter (Betreuer): Prof. Dr. rer. nat. Mark W. Greenlee

Gutachter: Prof. Dr. rer. biol. hum. Ulrich Frick

PREFACE

The aim of the thesis entitled “Assessment of Visual Literacy – Contributions of Eye Tracking” is to explore cognitive strategies in visual problem-solving in order to make a cognitive-psychological contribution to the assessment of Visual Literacy (VL). Solution strategies used by VL experts and novices are analyzed with eye tracking in combination with latent structure models.

The dissertation is arranged in three chapters: Introduction, Publications, and Conclusion. The references of all three publications are merged into one bibliography at the end of the thesis. All three papers were submitted to peer-reviewed journals (impact factor > 2.0) and entered their review process. Two out of the three projects have been successfully published. All publications are presented as pre-print versions of the published papers. The yet unpublished paper is represented in its most recently revised version. The manuscripts of all publications are formatted according to the guidelines of the American Psychological Association (APA). Layout of figures and tables have been adapted accordingly. No other changes have been made to the manuscripts.

The presented studies are part of the research project “Bildkompetenz in der Kulturellen Bildung” (BKKB) funded by the Federal Ministry of Education and Research (grant number: 01JK1606A). The overarching goal of BKKB was to lay groundwork for the development of an assessment battery that could measure VL. Contributions of co-authors are listed on page 3 and 4 and after each publication in chapter 2.

The dissertation has a special focus on the results from the eye-tracking experiments. Further analysis of VL assessment items and BKKB research results can be found at Frick, Rakoczy, Tallon, and Weiß (2020). For BKKB results on the influence of students’ social background on cultural aesthetic practice and motivation in art class see Weiß-Wittstadt (2022).

EIDESSTATTLICHE VERSICHERUNG

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Textpassagen, Daten, Bilder oder Grafiken sind unter Angabe der Quelle gekennzeichnet.

Bei der Auswahl und Auswertung des Materials haben mir die auf den Seiten 3-4 aufgeführten Personen in der jeweils detaillierten Weise unentgeltlich geholfen; dies ist auch in der Dissertation an den entsprechenden Stellen explizit ausgewiesen.

Weitere Personen waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- beziehungsweise Beratungsdiensten (Promotionsberater oder anderer Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen.

Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ich versichere an Eides Statt, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

Vor Aufnahme der obigen Versicherung an Eides Statt wurde ich über die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung belehrt

Regensburg, ,

DANKSAGUNG

Großen Dank möchte ich meinem Doktorvater Prof. Dr. Mark W. Greenlee für seine hervorragende fachliche Betreuung, seine Unterstützung und die offene Tür an der Universität Regensburg aussprechen. Am Lehrstuhl für Experimentelle Psychologie konnte ich an wichtigen forschungsnahen Lehrveranstaltungen teilnehmen. Insbesondere habe ich durch die Eye-Tracking Seminare von Prof. Dr. Kenneth Holmqvist wichtige Erkenntnisse für die Durchführung und Auswertung von Eye-Tracking Daten erwerben dürfen.

Ebenso möchte ich mich bei Prof. Dr. Ulrich Frick bedanken. Seine herausragende methodologische Expertise, seine Begeisterungsfähigkeit für interdisziplinäre Forschung und die kollegiale Unterstützung am Forschungszentrum der HSD Hochschule Döpfer in Köln war wesentlich für das Gelingen dieser Arbeit.

Auch möchte ich meinen Kolleginnen, Prof. Dr. Katrin Rakoczy und Dr. Susanne Weiß-Wittstadt aus dem BKKB Verbundprojekt für die intensive Zusammenarbeit während der Projektphasen und darüber hinaus danken. Die Erhebungen in den Schulklassen und unsere gemeinsamen Gruppentreffen werde ich immer in sehr guter Erinnerung behalten.

Weiterhin möchte ich dem ENViL Netzwerk und Prof. Dr. Ernst Wagner für den fachlichen Austausch zum Thema Bildkompetenz und Kunstdidaktik danken. Ich danke allen Expertinnen und Experten aus ENViL für die Teilnahme und die intensiven Diskussionen im Projekt. Ich danke allen Schulen, Lehrkräften und Schüler*innen für die Teilnahme an den BKKB Studien.

Abschließend möchte ich meiner Familie und meiner Freundin Sinem für ihre Geduld, ihr Verständnis und ihre Unterstützung von ganzem Herzen danken.

CONTENTS

Preface ... iii
Eidesstattliche Versicherung ... iv
Danksagung ... v
Contents ... 1
Abbreviations ... 2
Contributions ... 3
Abstract ... 5
Chapter 1. Introduction ... 6
Visual Literacy ... 6
Perceptual Learning ... 8
Visual Expertise ... 9
Research Questions ... 9
Methodological Approach ... 10
Eye Tracking ... 10
Latent Class and Latent Profile Analysis ... 12
Hidden Markov Models ... 13
Bradley Terry Models ... 16
Chapter 2. Publications ... 17
Preparations ... 17
Paper 1: Comprehension of Visual Logical Models ... 19
Paper 2: Visual Search on Artworks ... 47
Paper 3: Judgment of Visual Abstraction ... 79
Chapter 3. Conclusion ... 105
Summary ... 105
Limitations ... 108
Outlook and Concluding Remarks ... 110
Bibliography ... 113
Appendix A ... 140
Appendix B ... 144
Appendix C ... 148

Abbreviations

AIC	Akaike Information Criterion
AOI	Area of Interest
BIC	Bayesian Information Criterion
BKKB	Bildkompetenz in der Kulturellen Bildung (*Visual Literacy in Cultural Education*)
BPM	Business Process (Management) Model
BTM	Bradley-Terry(-Luce) Model
CEFR-VL	Common European Framework of Reference for Visual Literacy – Prototype
EMME	Eye Movement Modeling Examples
ENViL	European Network for Visual Literacy
HMM	Hidden Markov Model
IDT	Identification by Dispersion-Threshold
IVT	Identification by Velocity-Threshold
LC	Latent Class
LCA	Latent Class Analysis
LLBT	loglinear Bradley Terry models
LPA	Latent Profile Analysis
M	Mean value
MOB	model-based (recursive partitioning)
PC	paired comparison
PDF	probability density function
PM	Process Model
SD	Standard Deviation
SEM	Standard error of the mean
VL	Visual Literacy
WS	White Space
π	(1) LCA posterior membership probability, (2) initial state distribution probability for Hidden Markov Models, (3) worth parameter for Bradley Terry models
χ^2	Chi-square

CONTRIBUTIONS

The following people have contributed to this thesis.

Paper 1	**Comprehension of Business Process Models:** **Insight into Cognitive Strategies via Eye Tracking**
Authors	Miles Tallon[a, b], Michael Winter[f], Rüdiger Pryss[f], Katrin Rakoczy[d, g], Manfred Reichert[f] , Mark W. Greenlee[a], Ulrich Frick[b]
Contributions	MT, UF and KR designed the study. MT, MW, MR, and RP designed the process models, their presenting conditions, and constructed the related answering statements testing comprehension. MT and MW wrote the program, under supervision of MR and RP, and implemented it on the tablets. MT performed the field work. MT and MWG designed and analyzed all eye-tracking measurements. MT and UF performed all psychometric and other statistical analyses. MT wrote the first draft of the manuscript. UF, KR and MWG made editorial suggestions for the final version of the submitted manuscript. All authors read the manuscript and made suggestions for revisions. Kenneth Holmqvist commented on an earlier version of the submitted manuscript. Two anonymous reviewers evaluated the submitted manuscript. Their suggestions were incorporated into the final version of the published article.
Publication status	Published in Expert Systems with Applications, December 1_{st} 2019 https://doi.org/10.1016/j.eswa.2019.06.032

Paper 2	**How do Art Skills Influence Visual Search? –** **Eye movements Analyzed with Hidden Markov Models**
Authors	Miles Tallon [a,b] , Mark W. Greenlee[a], Ernst Wagner[c], Katrin Rakoczy[d,g], Ulrich Frick[b]
Contributions	MT, MWG, UF and KR designed the study. MT and EW selected and prepared the stimuli. MT conducted the field work. MT and UF designed and MT performed the statistical analysis. EW contributed aesthetic theory to the interpretation of statistical results. MT, MWG, KR and UF prepared the manuscript. All authors reviewed the manuscript and approved the submitted

	version. Two anonymous reviewers evaluated the submitted manuscript. Their suggestions were incorporated into the final version of the published article.
Publication status	Published in Frontiers in Psychology, January 28th 2021 https://doi.org/10.3389/fpsyg.2021.594248

Paper 3	**Assessing Heterogeneity in Students' Visual Judgment: Model-Based Partitioning of Image Rankings**
Authors	Miles Tallon[a, b], Mark W. Greenlee[a], Ernst Wagner[c], Katrin Rakoczy[d], Wolfgang Wiedermann[e], Ulrich Frick[b]
Contributions	MT, UF and KR designed the study. MT and EW selected and prepared the stimuli. MT conducted the field work and eye-tracking experiments. MT performed the statistical analysis with input on data interpretation by WW, UF and MWG. MT prepared the manuscript. All authors reviewed the article and approved the submitted version. Two anonymous reviewers evaluated the submitted manuscripts. Their suggestions will be incorporated into the final version of the article.
Publication status	Under peer-review in Frontiers in Psychology, May 2022 Published version available in Frontiers in Psychology, August 10th 2022 https://doi.org/10.3389/fpsyg.2022.881558

[a] Department of Experimental Psychology, University of Regensburg, Regensburg, 93053, Germany [miles.tallon@ur.de; mark.greenlee@ur.de]

[b] HSD Research Centre Cologne, HSD University of Applied Sciences, Cologne, 50676, Germany [m.tallon@hs-doepfer.de, u.frick@hs-doepfer.de]

[c] Academy of Fine Art Munich, Munich, 80799l, Germany [ernst@wagner-mchn.de]

[d] Institute for School Education and Empirical Educational Research, Justus-Liebig University, Gießen, 35394, Germany [katrin.rakoczy@erziehung.uni-giessen.de]

[e] Missouri Prevention Science Institute and Department of Educational, School, and Counseling Psychology, University of Missouri, Columbia [wiedermannw@missouri.edu]

[f] Institute of Databases and Information Systems, Ulm University, Ulm, Germany [michael.winter@uni-ulm.de, ruediger.pryss@uni-ulm.de, manfred.reichert@uni-ulm.de]

[g] formerly: DIPF Leibniz Institute for Research and Information in Education, Frankfurt am Main, 60323, Germany

Abstract

Visual Literacy (VL) is defined as a set of competencies to understand and express oneself through visual imagery. An expansive model, the Common European Framework of Reference for Visual Literacy (CEFR-VL) (Wagner & Schönau, 2016), comprises 16 sub-competencies, including abilities such as *analyzing*, *judging*, *experimenting with* or *aesthetically experiencing* images. To empirically assess VL sub-competencies different visual tasks were presented to VL experts and novices. Problem-solving behavior and cognitive strategies involved in visual logical reasoning (Paper 1), Visual Search (Paper 2), and judgments of visual abstraction (Paper 3) were investigated. Eye tracking in combination with innovative statistical methods were used to uncover latent variables during task performance and to assess the possible effects of differences in expertise level. Furthermore, the relationship between students' self-reported visual abilities and their performance on VL assessment tasks is systematically explored.

Results show how effects of perceptual skills of VL experts are less pronounced and more nuanced than implied by VL models. The comprehension of visual logical models does not seem to depend much on VL as experts and novices did not differ in their solution strategies and eye movement indicators (Paper 1). In contrast, the visual search task on artworks revealed how experts were able to detect target regions with higher efficiency than novices revealed by higher precision of fixations on target regions. Furthermore, latent image features were detected by experts with more certainty (Paper 2). The assessment of perceived level of visual abstraction revealed how, contrary to our expectations, experts did not outperform novices but despite that were able to detect nuanced level of abstraction compared to student groups. Distribution of fixations indicate how attention is directed towards more ambiguous images (Paper 3). Students can be classified based on different levels of visual logical comprehension (Paper 1), on self-reported visual skills, and the time spent on the tasks (Paper 2, Paper 3). Self-reported visual art abilities of students (e.g., imagination) influences the visual search and the judgment of visual abstraction.

Taken together the results show how VL skills are not determined solely by the number of correct responses, but rather by how visual tasks are solved and deconstructed; for example, experts are able to focus on less salient image regions during visual search and demonstrate a more nuanced interpretation of visual abstraction. Low-level perceptual abilities of experts and novices differ marginally, which is consistent with research on art expertise. Assessment of VL remains challenging, but new empirical methods are proposed to uncover the underlying components of VL.

Chapter 1. Introduction

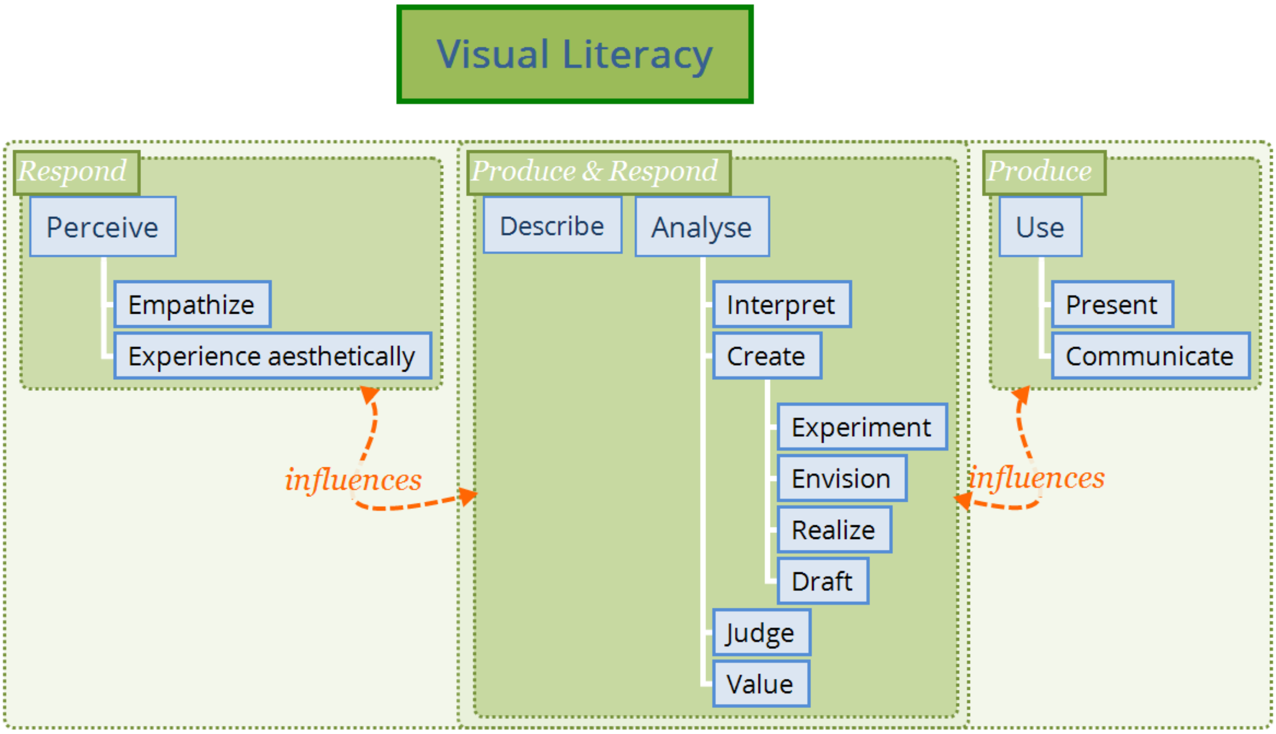
Visual Literacy

Visual Literacy (VL), also referred to as Visual Competency (Schönau & Kárpáti, 2019), is defined as a group of sub-competencies to actively decode visual messages, to discriminate, create, comprehend and appreciate images (Avgerinou & Pettersson, 2011; Brill et al., 2007). VL comprises multiple *sub-competencies* or *visual literacy competencies* (Loerts & Belcher, 2019; Tillmann, 2012). Most definitions refer to VL either as a *skill*, *ability* or *competency* often interchangeably (Avgerinou, 2003). The Common European Framework of Reference for Visual Literacy (CEFR-VL) builds upon the definition by Brill et al. (2007) and proposes a phenomenological model with 16 sub-competencies of VL (Wagner & Schönau, 2016). Basic competencies (*producing* and *responding* to images) are broken down into: analyze, communicate, create, describe, draft, empathize, envision, experience aesthetically, experiment with, interpret, judge, perceive, present, realize, use and value (Wagner & Schönau, 2016, p. 68). Figure 1 shows how most CEFR-VL sub-competencies include both producing and responding elements.

VL has been defined as a key competency of art education (Stokes, 2002; Vermeersch & Vandenbroucke, 2015). McMaster (2015) points out a significant research gap on how students create and deconstruct images and how VL is still an underdeveloped element of learning. What are the advantages of VL beyond better understanding of images? Current research supports the notion that visual language and communication play a key role in our cognitive and emotional development (Bentwich & Gilbey, 2017; Tyler & Likova, 2012) and can increase our ability to learn in other subject areas (Wagner, 2017), for example, to support reading comprehension (Kaya, 2020; Vaknin-Nusbaum & Nevo, 2021), to enrich informal learning environments (Guinibert, 2020), help to model and visualize data (Spalter & van Dam, 2008), or support science education (Güney, 2019).

Figure 1

The Common European Framework of Reference for Visual Literacy (CEFR-VL) with Sub-competencies Proposed by ENViL



Note. Adapted from Wagner and Schönau (2016). *Common European Framework of Reference for Visual Literacy-Prototype*. Waxmann, pp. 67–68.

Some efforts have been put forward to quantify VL (Brown & Lockyer, 2007; Groenendijk et al., 2020). However, empirical assessment of VL is challenging, as many heterogeneous phenomenological models, including different sub-competencies of VL have been proposed by art education research (Avgerinou & Pettersson, 2011; Kędra, 2018). The CEFR-VL model has a focus on art education with examples of potential assignments for art class. However, there is no empirical foundation on the connection of proposed sub-competencies of VL. Furthermore, the model incorporates sub-competencies, such as *perceiving images*, which are conceivable as perceptual learning abilities and others, such as *aesthetically experiencing images*, which may be addressed by empirical aesthetics and art expertise research. As the characterizations of VL competencies may differ depending on the research focus and discipline, the relationship of VL to perceptual learning and visual expertise is therefore clarified in the context of this thesis.

Perceptual Learning

Perceptual learning is the experience-driven improvement of our psychophysiological senses, including (but not limited to) vision (Fahle & Poggio, 2002; Harris, 2014). In contrast to *procedural* learning, which is defined as the improvement in performance for a given task after repeated training, perceptual learning modifies perception and behavior as a result of sensory experience and is generally thought to be independent of conscious forms of learning (Greenlee, 2014). However, empirical studies frequently use similar conditions to study both types of learning, making fundamental differences between the two difficult (Gold & Watanabe, 2010). Furthermore, the association with a reward may still be required for at least some forms of perceptual learning (Bourgeois et al., 2016).

The effects of perceptual learning are relatively permanent as subjects are able to maintain enhanced discriminatory ability even after not performing the learned task for years (Frank et al., 2018; W. Li & Gilbert, 2009; Seitz, 2017), for example, the skill to discriminate between specific horizontal and vertical target textures (Karni & Sagi, 1993). Moreover, these changes are not merely coincidental, but adaptive, and thus impose advantages such as enhanced responsiveness even to weak stimuli (Gold & Watanabe, 2010; Harris, 2014). The improved perceptual skills rarely interfere with each other and there is little learning transfer to new stimuli and tasks (W. Li & Gilbert, 2008; Seitz, 2017). There is also evidence for individual differences in perceptual learning (Greenlee et al., 2014; Muller-Gass et al., 2017; Yang et al., 2020), that is, differences between task-specific components of learning (e.g., learning context or number of repetitions), and subject-specific learning abilities (e.g., sensitivity to punishment or openness to experience) (Dale et al., 2021). Learning can also occur in early stages of vision, for example, the *Vernier acuity*, the ability to distinguish between the offset of one line to the left or right of another line can improve with training (Fahle & Edelman, 1993). Other learning experiences, such as the *oblique effect*, that is, humans' higher sensitivity for vertical and horizontal lines compared to oblique orientations, can also be found in empirical aesthetics research, for example, the preference for original Mondrian (1874-1944) compositions in contrast to oblique rotations of Mondrian art (Latto et al., 2000).

Both procedural learning as well as perceptual learning are involved in skill acquisition and are linked to the formation of professional (visual) expertise (Greenlee, 2014; Kellman & Garrigan, 2009). We could consider a person who learned to discriminate between shapes or colors a perceptual expert. However, the term "expert" usually refers to someone who works in a professional domain that necessitates higher, more cognitive, and top-down processes.

Visual Expertise

Expertise is described as extraordinary performance in a domain-specific subject, attained through training and measured by behavior analysis (Ericsson & Towne, 2010). Visual expertise is also domain specific, i.e., the perceptual skillset necessary is bound to a specific visually intensive task (Gegenfurtner & van Merriënboer, 2017) and can be observed in professional domains with the need for high visual acuity. Expert radiologists for example, are particularly skilled in visual awareness and illness characterization when inspecting medical images (Fox & Faulkner-Jones, 2017; Kundel et al., 2007; Wood, 1999). Chess experts can memorize chess board states and identify relevant positions in milliseconds (Charness et al., 2001; Sheridan & Reingold, 2014).

However, unlike research fields such as chess or medicine (Finan, 2002; Reingold & Sheridan, 2011), the study of art expertise and artistic experience is particularly challenging as artists are a very diverse group of people (Chamberlain, 2018; Kozbelt & Seeley, 2007). Regarding low-level perceptual abilities of expert artists, there is some evidence for enhanced facial information processes (Gartus et al., 2020; Hsiao et al., 2021; Solso, 2001) and visuo-spatial abilities (Calabrese & Marucci, 2006). Chamberlain et al. (2019) described artists as "experts in visual cognition" as art students differ from non-art students in their capacity to exercise top-down control over attentional processing but do not differ much in low-level visual processing.

The CEFR-VL model includes sub-competencies that are commonly found in artists, art experts, and art teachers, but also in those individuals who have an avid interest in the visual arts or professional experience in appraising, or even producing (e.g., drawing or painting), artwork. Art expertise may be regarded as a subset of VL since it requires training and understanding of domain-specific norms and practices for encoding and decoding visual information (Bauer, 2014). In the context of this thesis *VL experts* are defined as individuals with knowledge and experience in the professional domain of the visual arts, these include art educators, art designers, photographers, gallerists, art students, and freelance artists.

Research Questions

VL in its most basic form refers to the fundamentals of visual perceptual learning (Kappas & Olk, 2008). However, top-down processes of professionally acquired skills and aesthetic experience (Leder et al., 2014) are essential to consider someone a visual (art) expert (Chamberlain, 2018; Kozbelt & Seeley, 2007; Schabmann et al., 2016). In its current form, the CEFR-VL by ENViL encompasses low-level visual processing skills as well as top-down skills necessary for visual art expertise. As a result, and to make a cognitive-psychological contribution to the clarification of the

VL skillset, this thesis employs a set of visual tasks based on the CEFR-VL sub-competencies *interpreting*, *analyzing*, and *judging* to determine whether VL experts and novices differ in their ability to solve visual problems and their use of cognitive strategies.

The following overarching research questions are addressed:

1. How do VL experts (art educators, artists) and novices (laypeople unrelated to visual arts) differ in their cognitive strategies while solving visual tasks?
2. How can eye tracking in combination with other empirical methods be used to identify and explicate VL competencies in students, novices and experts?

Eye-tracking data were used as an external validation of the BKKB assessment tasks, i.e., VL experts should outperform novices on the visual tasks. As art expertise research and empirical VL research are in their infancy (Chamberlain, 2018; Matusiak, 2020) new empirical approaches are proposed to advance the field.

Methodological Approach

Eye Tracking

All presented publications in this thesis include the method of eye tracking as an essential part of the data analysis. Eye tracking is widely used in psychological and behavioral sciences (Holmqvist et al., 2011; Holmqvist & Andersson, 2017; Orquin & Holmqvist, 2019), including decision making (Boisvert & Bruce, 2016), diagnostics (Chan et al., 2018), marketing (Wedel & Pieters, 2017), or human computer interaction (Mason et al., 2017).

There are multiple ways to detect and classify eye movements, i.e., to classify distinct eye movement types (e.g., fixations, saccades) from raw data (Andersson et al., 2017). *Fixations* are periods where the eye is relatively still and the participant looks at a specific location on a screen (Holmqvist & Andersson, 2017, pp. 22–23). Fixation durations commonly last 200–300ms but can last multiple seconds depending on the task and situation. The eyes however are never completely motionless as intra-fixational micro-movements are always occurring (Martinez-Conde et al., 2013). By contrast, *saccades* are periods where the gaze shifts rapidly to another (target) position (Holmqvist & Andersson, 2017, p. 23). Besides the sampling frequency (Andersson et al., 2010), precision and accuracy of eye movement measurements are also influenced by detection algorithms and the mathematical model of the eye (Orquin & Holmqvist, 2018, 2019).

Two common event-detection algorithms are *Fixation Dispersion Algorithms* (Identification by Dispersion-Threshold, IDT) (Komogortsev et al., 2010) and *Velocity based algorithms* (Identification by Velocity Threshold, IVT) (Salvucci & Goldberg, 2000). IDT algorithms consider a fixation in data samples containing at least enough time to satisfy a set (or adaptive) duration and is located within a spatial area that does not exceed a dispersion threshold. Samples that meet these criteria are labeled as fixations. On the other hand, IVT algorithms identify fixations and saccades using a preset velocity threshold, where fixations are segments of samples with less than the set velocity threshold and saccades are segments with velocities greater than the given velocity threshold (Stuart et al., 2019). Combinations of both algorithm types, including other machine learning approaches, are possible and desirable for a valid event detection (Andersson et al., 2017; Zemblys et al., 2018). Micro-movements of the eye, e.g., micro-saccades (Engbert, 2006) are detectable but require higher sampling frequencies above 200Hz (Martinez-Conde et al., 2009).

Expertise-driven, top-down processes of eye movements often base their research on Yarbus's (Yarbus, 1967a) famous study on the effect of task instruction on eye movements: Yarbus (1967b) observed how the viewers' attention was drawn to the elements of the image that were relevant to the task at hand. Different task instructions before viewing the painting *They did not expect him* (Ilya Repin 1844-1930), such as "how long was the visitor gone?" or "how many people are in the room?" had significant influence on the individuals eye movement patterns. Recent research has confirmed the effect of task instruction on top-down eye movement control (Boisvert & Bruce, 2016; Borji & Itti, 2014; Haji-Abolhassani & Clark, 2014), for example, Simola et al. (2008) predicted different information search reading strategies based on the spatio-temporal distribution of eye movement patterns.

Eye tracking is also commonly used for research on expertise (Derek Panchuk et al., 2015; Fox & Faulkner-Jones, 2017; Gegenfurtner et al., 2011; R. Li et al., 2012; Tien et al., 2014; Vogt & Magnussen, 2007). Studies on visual expertise often list three major concepts that attempt to explain the superior visual abilities of experts over novices:

- Higher capacity of long-term working memory (Ericsson & Kintsch, 1995) to encode and retrieve visual information by experts;
- Selective attention to relevant stimuli of experts to reduce redundant information (Haider & Frensch, 1999); and
- An holistic approach to image perception that allows experts to extract more information from distal and para-foveal regions (Kundel et al., 2007).

A systematic review by Brams et al. (2019) emphasizes the importance of an efficient visual search rate, enhanced selective attention allocation, and an extended visual span for expert performance in sports and medicine. However, studies on art expertise are more rare and significant differences regarding number of fixations or fixation durations between art experts and novices are infrequent (Francuz et al., 2018; Vogt & Magnussen, 2007), for example, artists seem to fixate longer on familiar paintings while novices make longer fixations on unfamiliar paintings (Antes & Kristjanson, 1991). Instead differences are more often observed by scanpaths (fixations sequences) on task relevant high-level image features (Chamberlain et al., 2019; Jarodzka et al., 2010; Ylitalo et al., 2016). Artists seem to benefit from their holistic perception of images the most, as abstract, less salient regions of artwork are explored more (Koide et al., 2015; Kolodziej et al., 2018; Pihko et al., 2011; Vogt & Magnussen, 2007), and this may also help experts focus on the global overall structures during the creation of drawings (Drake et al., 2021).

Only recently has eye tracking been identified as an appealing methodology for VL research (Brumberger, 2021). Furthermore, the combination of eye tracking with advanced statistical methods may elevate the field of VL research, as it allows us to explore the intricate differences between experts and novices and the effects of VL on cognitive processes more closely. Details on the eye-tracking equipment and procedure used for this thesis can be found in Appendix A.

Latent Class and Latent Profile Analysis

Latent class analysis (LCA) and latent profile analysis (LPA) are statistical methods aimed to recover hidden groups from observed data (McCutcheon, 1987). LCA and LPA are used on continuous (LPA) or categorical (LCA) variables to detect the structure of relationships between subgroups to create typologies or scales (Oberski, 2016). Unlike factor analysis, which assumes that people differ by degrees on continuous latent dimensions, LCA/LPA models assume that people fall into latent categories. They are like clustering techniques, but more flexible because they can account for group uncertainty by using probability distributions for each class. LCA assumes *local independence* within each latent class (LC). The basic LC model can be formulated as

$$P(\boldsymbol{Y} = \boldsymbol{y}) = \sum_{x=1}^{C} P(X = x) \prod_{\ell=1}^{L} P(Y_\ell = y_\ell | X = x) \qquad (Eq.\ 1.1)$$

Where $P(\mathbf{Y}=\mathbf{y})$ is the probability for a complete response pattern, X represents the latent variable and $Y\ell$ the manifest variable, C is the total number of LCs, $P(X=x)$ indicates the proportion of individuals belonging to LC x. The probability of obtaining response pattern y, is a weighted average of the C class-specific probabilities $P(Y = y \mid X = x)$ (c.f. Vermunt et al., 2004). Typically, each person can be categorized to the LC with their highest posterior membership probability (Hagenaars & McCutcheon, 2002; Rost & Langeheine, 1997). This probability can be obtained by the Bayes rule (Vermunt et al., 2004);

$$P(\boldsymbol{X} = x|Y = y) = \frac{P(X = x)P(\boldsymbol{Y} = \boldsymbol{y}|X = x)}{P(\boldsymbol{Y} = \boldsymbol{y})} \qquad (Eq.\ 1.2)$$

In Paper 1 we used LCA to determine different classes for visual logical reasoning of students and adults. There we used the notation by Rost and Langeheine (1997), where π_g refers to the relative size of class g, and π_{ixg} refers to the posterior membership probability to belong to class g of G classes. In Paper 2 we applied LPA to describe different response profiles (solution patterns) during visual search.

Hidden Markov Models

Hidden Markov Models (HMM) are statistical methods to describe and analyze latent states and their transition probabilities over time (Rabiner, 1989; Visser et al., 2002). HMM have some important characteristics. Firstly, the states are discrete, as the data are sampled from many distributions with different parameters rather than following a single unimodal distribution. Secondly, the states are not directly observable (i.e., hidden) and can only be observed indirectly, as the mapping between states and observations is probabilistic rather than deterministic. Finally, the *Markov property*, i.e., every response is only dependent on the previous response, independent of all past states prior $t - 1$, creates an otherwise memoryless model beyond the most recent state transition (Visser, 2011).

The formal definition of a HMM is as follows (c.f. Rabiner, 1989)

$$\boldsymbol{\lambda} = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}) \qquad (Eq.\ 1.3)$$

A describes the transition probability array of state *j* following state *i*, where the state probability is independent of time *t*:

$$A = [a_{ij}], a_{ij} = P(q_t = s_j | q_{t-1} = s_i) \qquad (Eq.\ 1.4)$$

B describes the probability of observation k being produced from state j, independent of time t (sometimes called the *emission probability*):

$$B = [b_i(k)], b_i(k) = P(x_t = v_k | q_t = s_i) \qquad (Eq.\ 1.5)$$

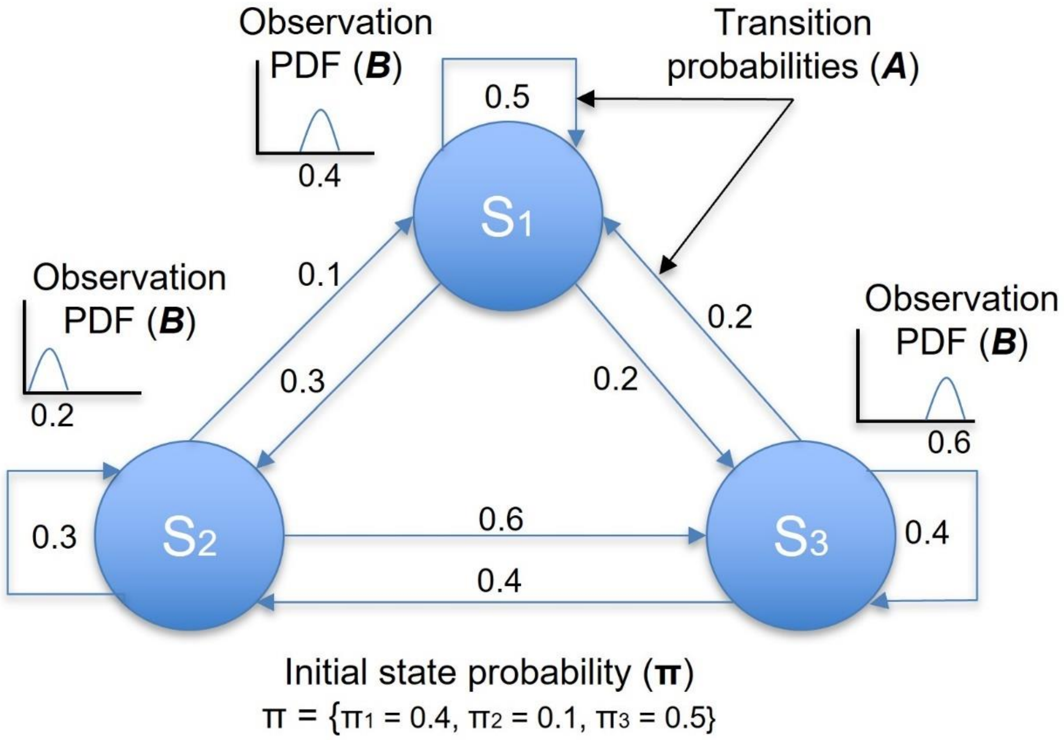
π describes the initial state probability:

$$\pi = [\pi_i], \pi_i = P(q_t = s_i) \qquad (Eq.\ 1.6)$$

A set of machine learning algorithms are used to reduce the complexity of calculations for HMM to compute the probability of the observation sequence effectively (evaluate), to discover the hidden state sequence that maximizes the probability of the observation sequence given the model parameters (decode), and to estimate the model parameters that best describe the model (training) (Visser, 2011). The algorithms that deal with evaluation, decoding, and training problems are called forward (backward), Viterbi and Baum-Welch algorithms, respectively (Rabiner, 1989). For a discussion on implementations see Blunsom (2004). Figure 2 shows an example of an HMM with three hidden states.

Figure 2

Hidden Markov Model with three Hidden States and Probability Distributions



Note. An HMM with three hidden states (S_1-S_3), transition probabilities (A), observation probabilities (B), and initial state distribution (π). The states are hidden to the observer and the output is a series of observations that are the outcomes of the observation probability density functions (PDFs). Adapted from "A computational model for task inference in visual search" by Haji-Abolhassani & Clark, 2013, Journal of Vision, 13(3):29, p. 6.

HMM have been popularized in speech recognition (Rabiner, 1989) but has since been used for a variety of applications (Mor et al., 2021) including eye movements, e.g., HMM as event-detection algorithms for eye movement types (Komogortsev et al., 2010). A study by Haji-Abolhassani and Clark (2013) used HMM to analyze visual search behavior on a grid with symbols and letters. The authors trained the HMM on a task modulated saliency map to infer between easy and difficult search processes given the observed eye movement patterns.

More recently HMM were also used to analyze eye movements during face perception (Chuk et al., 2014; Chuk et al., 2019; Hsiao et al., 2021). For eye movement data each fixation on an image can be interpreted as an observation following a probability distribution of an underlying image area (hidden state). The general concept that eye movement events are random variables manifested as the observable outputs of underlying stochastic processes is central to probabilistic models such as HMM (Boccignone, 2019; Coutrot et al., 2018).

Paper 2 goes into detail on how HMM can be used to analyze eye movement fixation sequences during visual search on artworks.

Bradley Terry Models

Objects or their properties can be characterized by a systematic comparison in form of a paired comparison (PC) task. The PC is a psychometric scaling method whose basic principle was developed by Thurstone (1927). PC tasks are the most common method when objects are to be placed into an ordered ranking. PC tasks are especially useful when the items can only be evaluated subjectively (David, 1988). The statistical model by Bradley and Terry (1952) and Luce (1959), called the Bradley-Terry-Luce model (BTM), is frequently used to analyze such preference decisions.

The probability to prefer item i over j is defined by

$$P_{i>j} = \frac{\pi_i}{\pi_i + \pi_j} \qquad \text{(Eq. 1.7)}$$

where π_k is a strength parameter (also called *worth parameter*) for item k, $1 \leq k \leq K$ (c.f. Bradley & Terry, 1952). The worth parameters (π) indicate how likely an item is selected in a PC. For all comparisons, $p_{ij} + p_{ji} = 1$ holds, since a decision is forced and no draw is allowed.

Such forced-choice questioning techniques overcome further problems such as *end-aversion bias*, that is, the tendency to choose the middle of the scale closer to a neutral position (Choi & Pak, 2005). Furthermore, the assumption of equidistant response categories is not required for BT models. Even though BT models have been widely used in studies on comparison data, for example, to determine the best sport teams (Tutz & Schauberger, 2015) or to detect the individual preference for fashion models (Strobl et al., 2011), the presented application may also be useful for art-class assignments, as the individual sensitivity for latent images characteristics are made observable and quantifiable.

Paper 3 describes how BT models can be combined with model-based partitioning (MOB) to uncover heterogeneity in students’ judgments of visual abstraction.

CHAPTER 2. PUBLICATIONS

Preparations

All presented papers include data from two separate samples: a student sample, comprising high school students and an eye-tracking sample comprising VL expert and novice participants. Members of ENViL were invited to contribute to a generic framework for visual assessment tasks and to take part in pre-tests on selected assignments and eye-tracking items. All items were specifically programmed for the assessment tool and were presented on Android tablet screens (see Andrews et al., 2018). Data acquisition in students was conducted in a classroom setting in schools in Germany. School classes were recruited and informed on the details of the study by using printed or identical electronic leaflets that had been sent to school principals (see Appendix C). Up to 30 students were able to take part simultaneously.

The eye-tracking experiments were conducted with experts and novices on a selected group of assessment items. Participants were classified as VL-experts if they were members of the European Network of Visual Literacy (ENViL) or worked in professions that required a high level of visual competence (photographer, gallerist, art educator, art designer, art students, or self-employed artists). Novices in VL were recruited adults from various educational settings' clerical and academic personnel who stated that they were not particularly talented or familiar with visual arts or design. Multiple locations were allocated for eye-tracking recording sessions. These included a laboratory rooms at the HSD University of Applied Sciences in Cologne, Ulm University, and a seminar room at the Academy of Fine Arts in Munich (see Appendix A).

The presented publications focus on three different item-sets included in the VL assessment battery.

1) Understanding Business Process Models (BPM)

   To *interpret* images and objects as defined by the CEFR-VL requires one to assign meaning to them, i.e., to translate the effect of an image into words by reasoning after reflection on the basis of observation and knowledge of codes and conventions (Wagner & Schönau, 2016, p. 75). Similar to flowcharts, BPMs represent logical sequences of information and workflows in a visual form. Comprehension of BPMs can be seen as a special form of diagrammatic reasoning (Kazmierczak, 2001), where thought processes are explained through visual imagery instead of verbal or mathematical means. The CEFR-VL model does not specify how VL may influence the understanding of visual logical representations. The article explores if VL-experts benefit from their expertise when logical processes are presented in a visual form.

2) Visual Search on Artwork

The CEFR-VL considers *analyzing*, i.e., to attentively and accurately focus on visual stimuli and to identify characteristics of images (Wagner & Schönau, 2016, p. 70) to be an essential part of VL as it is closely related to multiple other sub-competencies (e.g., *perceiving*, *interpreting*). Continuous involvement in art and images by VL experts may influence cognitive strategies involved in visual search on images of artwork. *Analyzing* should therefore be crucial in visual search. The study reveals and visualizes efficient search behavior for identifying latent image features in artworks.

3) Judgment of Visual Abstraction

The CEFR-VL defines *judging* (or evaluating) images as the ability to formulate a justified statement or estimation about images and artistic creations (Wagner & Schönau, 2016, p. 76). Similar to the aesthetic appreciation of art (Leder et al., 2012), the judgment of artwork may be affected by specific image features as well as individual characteristics of viewers. Therefore, it is interesting to see how the individual judgments of VL experts and novices differ when tasked to rank images by their level of visual abstraction. The presented approach in paper 3 makes the underlying preference judgment of visual abstraction quantifiable.

Paper 1: Comprehension of Visual Logical Models

Comprehension of Business Process Models: Insight into Cognitive Strategies via Eye Tracking

Miles Tallon, Michael Winter, Rüdiger Pryss, Katrin Rakoczy, Manfred Reichert, Mark W. Greenlee, Ulrich Frick

Author-produced version of this article published after peer-review on December 1st 2019 in Expert Systems with Applications https://doi.org/10.1016/j.eswa.2019.06.032. (Tallon et al., 2019) reprinted with permission.

Abstract

Process Models (PM) are visual documentations of the business processes within or across enterprises. Activities (tasks) are arranged together into a model (i.e., similar to flowcharts). This study aimed at understanding the underlying structure of PM comprehension. Though standards for describing PM have been defined, the cognitive work load they evoke, their structure, and the efficacy of information transmission are only partially understood. Two studies were conducted to better differentiate the concept of *visual literacy* (VL) and *logical reasoning* in interpreting PM.

Study I: A total of 1047 students from 52 school classes were assessed. Three different process models of increasing complexity were presented on tablets. Additionally, written labels of the models' elements were randomly allocated to scholars in a 3-group between-subjects design. Comprehension of process models was assessed by a series of 3*4 (=12) dichotomous test items. Latent Class Analysis of solved items revealed 6 qualitatively differing solution patterns, suggesting that a single test score is insufficient to reflect participants' performance.

Study II: Overall, 21 experts and 15 novices with respect to visual literacy were presented the same set of PMs as in Study I, while wearing eye-tracking glasses. The fixation duration on relevant parts of the PM and on questions were recorded, as well as the total time needed to solve all 12 test items. The number of gaze transitions between process model and comprehension questions was measured as well. Being an expert in visual literacy did not alter the capability of correctly understanding graphical logical PMs. Presenting PMs that are labelled by single letters had a significant influence on reducing the time spent on irrelevant model parts but did not affect the fixation duration on relevant areas of interest.

Both samples' participants required longer response times with increasing model complexity. The number of toggles (i.e., gaze transitions between model and statement area of interest) was predictive for membership in one of the latent classes. Contrary to expectations, denoting the PM events and decisions not with real-world descriptions, but with single letters, led to lower cognitive workload in responding to comprehension questions and to better results. Visual Literacy experts could neither outperform novices nor high-school students in comprehending PM.

Keywords: visual literacy; business process model; eye tracking; latent class analysis; cognitive workload

1 Introduction

What are process models?

A process model (PM) is a textual or visual representation, which documents all steps of an entire process (Schultheiss & Heiliger, 1963). Thereby, visual process models, *inter alia*, allow the depiction of complex algorithms, business steps, or logistical operations in a descriptive form (Aguilar-Savén, 2004; Bharathi et al., 2008; Rojas et al., 2016). PM should be designed such that practitioners can apply them for their tasks at hand (Roehm et al., 2012; Ungan, 2006). Moreover, PMs have to be understandable by all practitioners (Reggio et al., 2015; Zimoch, Pryss, et al., 2017). Existing research on process model comprehension has considered two groups of factors: (1) Subjective capability (e.g., model reader expertise) should be distinguished from (2) objective characteristics of the model itself (e.g., process model complexity).

For objective factors, a framework has been proposed (Moody et al., 2002) to evaluate the quality of process models. Notational deficiencies (e.g., semantic transparency) and their influence on the comprehension of process models have been reported by Figl et al. (2013). Regarding subjective factors, Recker and Dreiling (2007) compared two popular process modeling languages (business process model notation BPMN and event-driven process chain EPC). These studies focus on subjective aspects of PM comprehension, since they conclude that subjective factors have a greater impact than objective factors. A recent overview on studies investigating subjective as well as objective factors of PM comprehension is provided by (Figl, 2017).

Understanding PMs may not only be regarded as an endpoint depending on both factors described above, but also as a key competence for a multitude of cognitive tasks that share in common the classification and ordering of events and decisions into meaningful sequences (Dumas et al., 2013). As PMs are mostly presented as charts following specific rules of formalization in a standardized notation, it seems to be of interest to analyse the interplay between the visual inspection of charts representing PMs and their comprehension (Dumas et al., 2012).

Semantic Notation of PM

After a series of experiments with both subjective (i.e., cognitive load, Sweller et al. (2011) and objective factors (i.e., semiotic theory), Mendling et al. (2012) conclude that additional semantic information impedes syntax comprehension, whereas theoretical knowledge facilitates syntax comprehension.

The study at hand tries to open up the perspective of PM comprehension from pure graphical notation to semantic notions (real-world problem descriptions versus symbolic notation) as well as to personal capacities necessary for model comprehension (psychometric measurement of competence types or levels). Recker and Dreiling (2011) also highlight the importance of understanding subjective factors to enable development of understandable PMs.

Visual Literacy

Subjective factors play a key role in the understanding of PMs. It is therefore of interest to take a closer look at the ability of attentively analysing and interpreting images, an ability that is coined as Visual Literacy (VL; see Avgerinou and Pettersson (2011)). From the review by Figl (2017), it becomes clear that the construct of VL has not yet been used to analyse potential interactions between subjective and objective factors with respect to model comprehension. To the best of our knowledge, with the exception of a recent study (Bačić & Fadlalla, 2016), whose authors focused more on visual *intelligence* than on *literacy*, no study has yet been published dealing with the concept of Visual Literacy and its impact on PM comprehension. This is even more astonishing considering that VL has been postulated as a basic competence underlying the precise deciphering of images (receptive component of VL), the production of such images, as well as the reflection on the constituent processes (Wagner & Schönau, 2016). Images guide our perception of the world, our preferences, and our decisions, and VL is considered a central goal of arts education (Wagner and Schönau, 2016). Whether or not a good capability of analysing, memorizing, and envisaging visual stimuli is helpful for the comprehension or production of PMs (Brumberger, 2011), has yet to be determined.

It also remains unclear whether VL can be measured like an IQ score on a continuum of homogeneous tasks representing the same, continuously distributed latent trait, best assessed by a “Rasch scale” (see Boy et al. (2014) for an example in the field of visualization capability). By contrast, VL might also represent a categorical model (Brill et al., 2007), for which different groups of people have specific gifts and talents in common, qualitatively differing from each other without the possibility of representing these differences by a single score (latent class model, see McCutcheon (1987).

Eye tracking as measurement for PM comprehension

Eye-tracking methods help to understand and visualize underlying cognitive processes in problem solving (Bednarik & Tukiainen, 2006). Thus, eye-tracking can help to externally validate the measurement method of VL. Eye-tracking has been established in the investigation of competence and competence acquisition (Jarodzka et al., 2017). Conclusions about strategies or procedural knowledge can be drawn by analysing the processing of visual tasks that, otherwise, could not have been verbalized or could only be partially verbalized by the subjects retrospectively (Reingold & Sheridan, 2011; Sheridan & Reingold, 2014). The underlying cognitive processes thus may be better understood (Lai et al., 2013). Eye-tracking measures have provided insights into differences in experts and novices (Gegenfurtner et al., 2011; Vogt & Magnussen, 2007), the prediction of fluid intelligence (Laurence et al., 2018), as well as distinguishing between strategies in spatial problem solving (Y.-C. Chen & Yang, 2014).

PM comprehension has been studied by means of eye tracking (Figl, 2017; Hogrebe et al., 2011; Petrusel & Mendling, 2013; Zimoch et al., 2018; Zimoch, Mohring, et al., 2017), but not from the viewpoint of VL. It could be shown that subjects providing correct responses to comprehension questions after regarding a graphical model had fixated longer on relevant parts of the respective PM than on irrelevant parts (Petrusel & Mendling, 2013; Zimoch et al., 2018).

Cognitive strategies analysed by eye movements have been studied for graphically oriented intelligence tests (Hayes et al., 2011; Vakil & Lifshitz-Zehavi, 2012). A recent study by (Laurence et al., 2018) could predict from eye movement indicators approximately 45% of the variance of "Wiener Matrizen Test 2" (Formann et al., 2011) test results. Toggling (gaze transition between two areas of interest) has been shown to be the most reliable measure (Laurence et al., 2018) in this context. Other typical measurements include pupillometry (van der Meer et al., 2010) or fixation distribution (Najemnik & Geisler, 2005); (Bucher & Schumacher, 2006). Based on previous results on the analysis of matrix-based cognitive tests, the present study enhances the spectrum of visual tasks and tries to compare similar output measures for the comprehension of PMs.

To conclude, this study contributes to further analysing comprehension of PMs by using eye-tracking data. Previous studies have shown that experts in their professional domain (e.g. art, medicine, chess) fixate longer on task relevant parts and shorter on task redundant parts (Gegenfurtner et al., 2011). It has yet to be determined how the comprehension of graphically presented logical models is influenced by VL.

Research goals and objectives

This study aims to apply psychometric concepts to the field of PM research. Moreover, we try to corroborate these efforts by using innovative technology (i.e., eye-tracking measurements).

Notably, the role of expertise in VL for solving visual tasks seems unclear, and even questionable for comprehending PMs.

Based on the previous research on process model comprehension, this paper wants to contribute empirically to the influences on process model comprehension. Methodologically, this is accomplished by means of (1) latent class analysis (LCA) and (2) eye tracking. Through LCA, we are able to determine if the answers given by students follow a homogeneous latent trait or should better be interpreted as qualitatively differing solution patterns. The use of eye tracking helps to identify potential differences in participants' understanding by analysing where and for how long subjects fixate PM aspects. Cognitive load theory (Sweller et al., 2011) interprets these measurements as indicators for cognitive workload.

In summary, three major research questions are addressed in this paper:

1) How can the comprehension of PMs be measured in a population of students? More specifically, do answering patterns follow a homogeneous latent trait or should they be interpreted as qualitatively differing solution patterns?

2) How do features of PMs have an impact on the general PM comprehension?
   a. Do students successfully decipher the graphical notation (e.g., logical symbols like arrows, "x" or "+")?
   b. How does the semantic notation of PMs influence the response time and the PM comprehension?
   c. What effect does the model complexity have on response time and comprehension?

3) How does the competence level in analysing and interpreting images (VL) covary with PM comprehension?
   a. How do VL experts and novices differ in fixation duration on relevant rsp. redundant parts of the PMs?
   b. How does the expertise in VL covary with the eye movement's volatility of gaze transitions?

2 Materials and Methods

Subjects

Sample I comprised 1047 high-school students from 52 classes (9th to 13th grade: 21, 28, 1, 1, 1) in 29 schools in Germany. Overall, 52.5% were female, the average age was 15.27 years (SD = 0.94). Schools were recruited in the federal states of Hessen, North-Rhine Westphalia, Schleswig-Holstein, and Rhineland Palatinate via leaflets, letters and personal recommendations. The test was conducted in regular classrooms. Up to 30 students were able to participate in the test simultaneously. In Sample I understanding PM was one segment of a longer (duration: 45 minutes) test on Visual Literacy. All answers were given via touchscreen input by the participants. School classes were offered a lump sum of 100€ as collective compensation.

Participants in Sample II were enrolled as experts in visual literacy (n=21), if they were members of the European Network of Visual Literacy (ENViL) or working in professions requiring a high visual competence (photographer, gallerist, art educator, art designer, art students, or self-employed artists). Novices (n=15) in visual literacy were adults from the clerical and academic staff of various educational settings declaring themselves as not overwhelmingly talented or familiar with arts and visual design. The age span ranged from 16 to 66 years (M = 29.5). All participants had normal or corrected-to-normal vision. Student participants in Sample II received 20€ each as compensation. Other participants, including the expert group, who were intrinsically interested in the topic of Visual Literacy and eye tracking, participated without further compensation.

The study was conducted according to the guidelines for human research outlined by the Declaration of Helsinki and was approved by the Ethics Committee of Research of the Leibniz Institute for Research and Information in Education (DIPF, 01JK1606A). All subjects (and their legal representatives respectively) had given written informed consent.

Materials and procedure

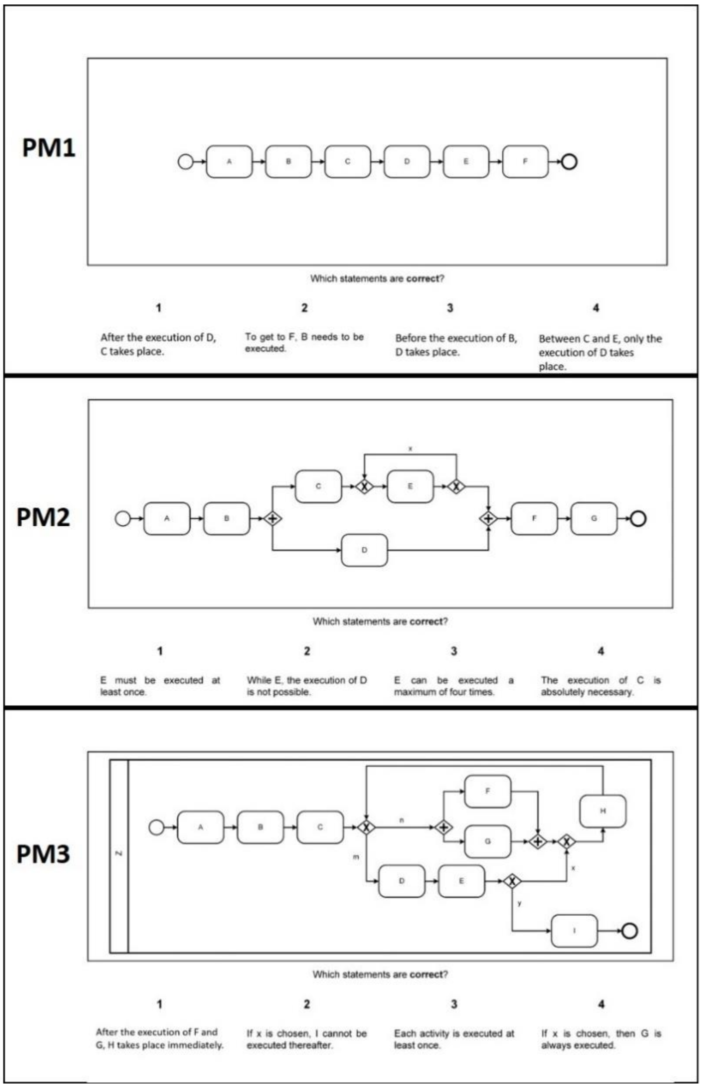
The assessment in both samples was conducted on Android A6 Tablets with 10.1-inch screen size. All test items were programmed specifically for the assessment tool (Andrews et al., 2018). The process models were created in BPMN 2.0 (OMG, Object Management Group, 2011). This language serves as an industry standard and constitutes the most widely used process modeling language (Allweyer, 2016).

All participants were given the identical instruction on the tablet screen: "In the following, different processes are presented in the form of process models. A process model visualizes the sequence of events and decisions. Try to understand the process in the process model and select all correct statements (multiple statements can be correct)."

Participants were required to inspect three subsequently presented PMs and to evaluate 4 statements based on the respective model, thereby representing a within-subject factor with three factor levels (Fig. 1). Statements were balanced for affirmation and rejection to indicate the correct response. The models were ordered in increasing complexity, where each new model included more activities (boxes) and gateways (inclusive, exclusive or parallel paths). Furthermore, in order to ensure a proper increase in process model complexity, the process models were created using the guidelines from Becker et al. (2000) and the adopted cognitive complexity measure proposed in Gruhn and Laue (2006). The comprehension statements as well as the activity-labels in the respective "boxes" of each process model were randomly allocated to each subject in one of three different verbal frames, thereby representing a between-subjects factor with the following factor levels: Letters (L), Sentences (S) and Pseudo Sentences (P). This manipulation means that events in the process models as well as in the comprehension test items were either denoted with a single letter (e.g. "execute F"), a meaningful sentence describing an everyday situation (e.g. "read Facebook message"), or with a pseudo sentence (e.g. "An ecap with mistives cannot be handed over") using meaningless artificial nouns to describe the events.

Figure 1

Process Models (PM1, PM2, PM3) in the Letter Condition.



Note. PMs were presented to respondents in increasing complexity. The boxes (activities) include actions to be performed, the arrows (sequence flow) define the execution order of activities, the x (an exclusive gateway) splits the routes of the sequence flow to exactly one of the outgoing branches. The + symbolizes a parallel gateway that is used to activate all outgoing branches simultaneously.

For Sample II, SMI eye-tracking glasses were used (SMI ETG 2w Analysis Pro). The glasses were positioned onto the subject's head, and the subjects were free to move their heads during task completion. Subjects were seated 50-80 cm away from the tablet screen. All eye-tracking data were recorded at 60Hz. Saccades and fixations (as well as blinks) were recorded binocularly and computed by the SMI event detection algorithm. Each session started with a 3-point calibration following the standard procedures for SMI iViewTM. The default eye movement parameters from SMI BeGazeTM version 3.7 were used. A fixation cross was displayed between each trial for 2 seconds. More details of the procedure and on data processing for eye-tracking measurements are given in a supplementary e-appendix (Appendix A).

Measurement and Data Analysis

The vector of 12 responses given on the tablets was transformed into 12 dichotomous items *x* representing each a correct judgement of the underlying verbal statement (1 = correct). The vector ***x*** of judgements then was analysed by latent class models (Dayton & Macready, 2006) describing typical solution patterns among the participants.

$$p(\boldsymbol{x}_v) = \sum_{g=1}^{G} \pi_g \prod_{i=1}^{k} \pi_{ixg} \qquad \text{where:} \qquad \sum_{g=1}^{G} \pi_g = 1 \qquad (1)$$

with *g* := number of latent class (*1 .. G*), *x* := response chosen on item *i (1 .. k)*, xν vector of correct judgments, π_g := relative size of class *g*, and π_{ixg} probability of choosing response *x* on item *i* given class *g*. Model parameters (π_g, $\pi_{ix|g}$) were estimated with MPLUS (6.0) software for all LCA solutions between 2 and 8 latent classes. The best number of latent classes was decided on model fit criteria (AIC, BIC) and the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test, as well as the Lo-Mendell-Rubin adjusted LR test implemented in MPLUS (Asparouhov & Muthén, 2012). In order to prevent local maxima of the likelihood function of the estimated parameters, the number of initial stage random starts was set to 1000, and the number of final stage optimizations to 50 for each number of classes. The estimated model parameters (π_g, $\pi_{ix|g}$) can be used to calculate membership probabilities for each participant in every latent class g in the following way (see equation 37, Rost and Langeheine (1997) p. 29).

$$p(g|x_v) = \frac{\pi_g \prod_{i=1}^{k} \pi_{ixg}}{\sum_{h=1}^{G} \pi_h \prod_{i=1}^{k} \pi_{ixh}} \qquad (2)$$

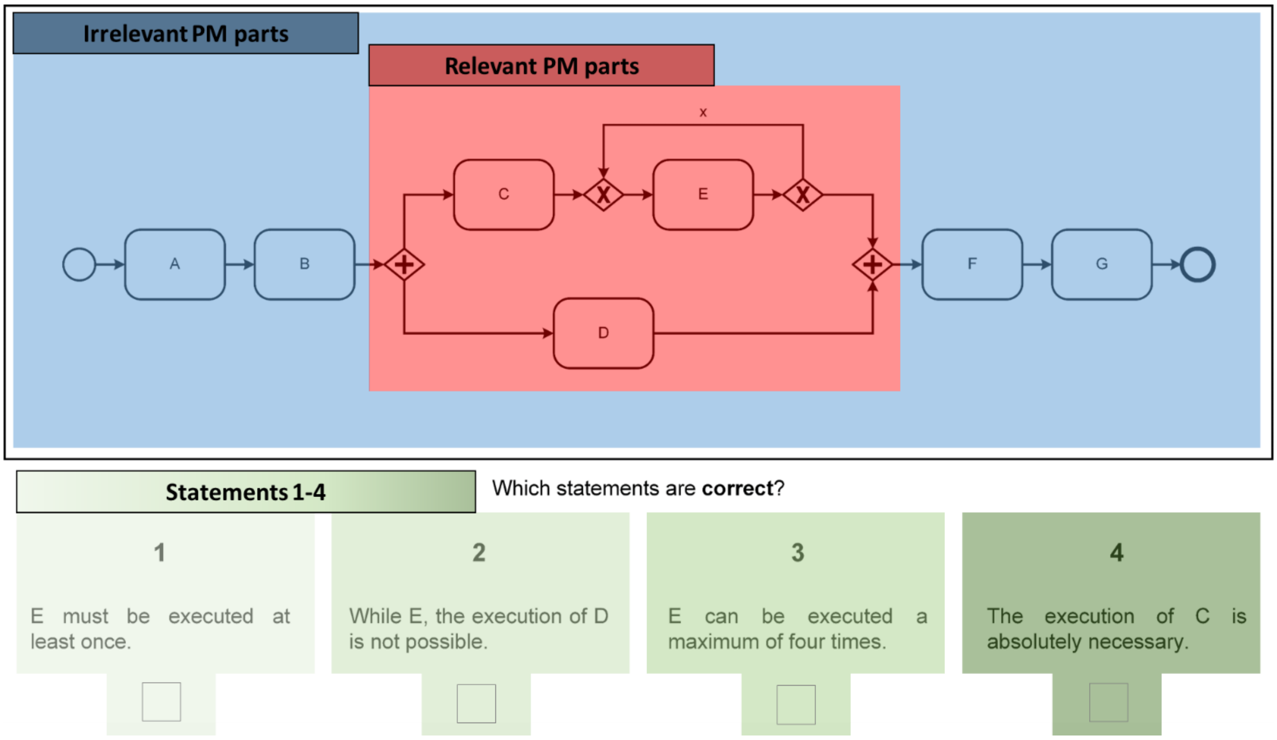
Based on the modal value, each participant was classified in his/her most probable latent class. Participants from Sample II were also classified using their response patterns and the item parameters estimated from Sample I. Additional measurements in Sample II were based on the following eye-tracking characteristics: a) response latency, which is the time spent on each trial in seconds, b) fixation duration on PM, which is the sum of all fixation durations on the model, c) fixation time on statements, which is the time spent on fixating the four response statements, d) number of toggles, which is the number of transitions between model and responses, and e) toggling rate, which is the number of toggles between model and responses divided by response latency.

Transitions between model and responses were counted each time the subject's gaze moved from model area of interest (AOI) to any statement AOI or vice versa. Whenever the gaze would stop to fixate on regions that were not defined by any AOI ("White Space"), the transition was not counted as a toggle.

Fixations for each trial were mapped on corresponding reference images by a single rater (MT) using SMI fixation-by-fixation semantic gaze mapping. For a comparison to frame-by-frame mapping see (Vansteenkiste et al., 2015). Independent ratings were performed (by MW) based on complete datasets of two randomly chosen subjects. In our study we reached a high inter-rater-reliability (Cohen's Kappa > 0.94 for all PMs). Figure 2 shows the AOIs of the second PM. Relevant parts of the graphical model (coloured in red) that were necessary for correctly accepting/rejecting a statement were a priori determined by process modeling experts from Ulm University (Zimoch, Pryss, et al., 2017). The wording of all test items (in German) was also a result of expert discussions within the same group. All gaze data was acquired by SMI iView ETGTM software. The analyses were carried out with SMI eye-tracking software "BeGaze 3.7". Further information on the eye-tracking equipment, technical settings and calibration procedure can be found in the e-appendix of this article (Appendix A).

Figure 2

AOI Distribution for Process Model 2 (Parallel Paths, 1 Loop)



Note. Colors indicate irrelevant PM parts (blue), relevant PM parts (red), and relevant parts of answers 1-4 (green).

Differences between PMs were analysed using repeated measurement ANOVA models for all eye movement indicators. Due to the relatively small sample size, differences between groups of respondents on the same indicators (e.g. status of expertise) were tested using univariate GLM models. In order to test significant associations between latent class membership and eye movement indicators, dummy variables for the larger groups (LC4, LC5, and LC6, see section 3.2) were constructed. In separate models, response latency, fixation duration on redundant or relevant parts of PM2 (second model in order of appearance), fixation duration on response statements, and number of toggles between PM2 and answering statements were tested as predictors of class membership via logistic regression models. All subjects not classified into one of the three larger groups were incorporated as part of the respective reference group, against which the impact of, for example, toggles was tested to predict membership. Again, due to small sample size these calculations were performed only in univariate analyses (only one predictor) omitting multivariate relationships and interaction effects during these explorative analyses. All statistical tests beyond the experimental variation of conditions are regarded as purely explorative and therefore not subject to measures against inflation of Type-I error risk.

3 Results

Solution Patterns in Scholars in Sample I

Both criteria (AIC and BIC) displayed substantial improvement of model fit until the introduction of a sixth latent class to be estimated. A seventh class resulted in deterioration of the BIC index, and no statistically significant differences could be demonstrated compared to the more parsimonious model with 6 latent classes in both the Vuong-Lo-Mendell-Rubin Likelihood Ratio Test, and the Lo-Mendell-Rubin adjusted LR test. Therefore, six latent classes were chosen as the final solution.

Table 1 gives an overview on the item parameters $\pi_{ix|g}$, which denote the probability of a correct solution in each of the six latent classes for each comprehension item.

Table 1

Process Model Complexity and Latent Class Parameters in Sample I

Model complexity	Item	Test items (wording for Letter condition*)	Solution	Probability of correct solution in latent class (corresponds to π_{ixg} in formula 1)						
				LC1	LC2	LC3	LC4	LC5	LC6	Total Sample I
linear model	Q1	After the execution of D, C takes place.	reject	0.727	0.611	0.906	0.896	0.979	0.998	0.889
	Q2	To get to F, B needs to be executed.	accept	0.566	0.578	0.632	0.652	0.876	0.918	0.757
	Q3	Before the execution of B, D takes place.	reject	0.594	0.394	0.726	0.822	1	1	0.825
	Q4	Between C and E, only the execution of D takes place.	accept	0.337	0.351	0.34	0.489	0.766	0.576	0.541
parallel paths, 1 loop	Q1	E must be executed at least once.	accept	0	1	0	0	0.094	1	0.377
	Q2	While E, the execution of D is not possible.	reject	0	0.815	1	1	0	0.871	0.527
	Q3	E can be executed a maximum of four times.	reject	0.872	0.836	0	0.933	0.932	0.948	0.822
	Q4	The execution of C is absolutely necessary.	accept	0.126	0.208	0	1	0.064	0.436	0.287
linear, exclusive and inclusive gateways, 2 loops	Q1	After the execution of F and G, H takes place immediately.	accept	0.306	0.565	0.217	0.385	0.598	0.55	0.481
	Q2	If x is chosen, I cannot be executed thereafter	accept	0.244	0.323	0.396	0.341	0.597	0.548	0.456
	Q3	Each activity is executed at least once.	reject	0.448	0.497	0.604	0.644	0.78	0.608	0.630
	Q4	If x is chosen, then G is always executed.	reject	0.538	0.579	0.736	0.696	0.877	0.72	0.726

Note. Table 1 gives model parameters for all conditions. Red-shaded cells depict below-average probabilities (> |10%|) of solutions for the respective item in each latent class. Green-shaded cells signify above-average probabilities (> 10%) of correctly solved items.

Interpretation of latent class 1 (LC1) and latent class 6 (LC6) seems straightforward: LC1 represents a group of persons with rather poor chances to solve each of the comprehension items. Members display probabilities at least 10% below the chance rates of the whole sample. This group comprised about 13% of the sample and was called “under performers”. On the contrary, LC6 consists of about 31% of the participants with excellent performance: members had no comprehension probability below sample average, but most items were solved with slightly or clearly better (green cells: > 10%) probabilities than the total sample. LC6 were called “logic champions”.

LC2 (24%) closely resembles LC1 except that participants are most likely able to respond correctly to items 1 and 2 of the "parallel paths – 1 loop" model (PM2), which had zero probability in LC1. On the other hand, the group LC5 (10%) is quite similar to the largest group "logic champions" class (LC6), but it fails to recognize the correct solutions for question 1, 2 and 4 of the "parallel paths – 1 loop" model (PM2). LC2 can be labelled as "under-performers with understanding of simultaneous tasks", and LC5 as "logically correct thinking with misinterpretation of parallel paths".

LC3 represents a typical response pattern (12%) that is performing at an average level for all test items requiring a comparison of not more than two activities. But when 3 or more information units have to be combined for a correct solution, LC3 strongly underperforms (e.g. "After the execution of D, C takes place" (PM1,Q1) vs. "After the execution of F and G, H takes place immediately" (PM3, Q1). Therefore they were called "binary thinking group". Finally, the solution probabilities in LC4 (size 10%) display an excellent understanding of parallel paths (but misunderstand the "x" notation of loops), and a slightly below average comprehension of PM1 and PM3. Accordingly, this group was therefore called "multi-tasking group".

Table 2

Number of Latent Class Members by Model Condition in Sample I

Condition		**Latent Class**						
		1	**2**	**3**	**4**	**5**	**6**	**N Total**
Letter (L)	Frequency	35	56	20	24	11	191	337 (32.19%)
	Row %	10.39	16.62	5.93	7.12	3.26	56.68	
	Column %	25.93	22.13	16.26	22.64	10.58	58.59	
Sentence (S)	Frequency	49	121	46	33	63	42	354 (33.81%)
	Row %	13.84	34.18	12.99	9.32	17.8	11.86	
	Column %	36.3	47.83	37.4	31.13	60.58	12.88	
Pseudo Sentence (P)	Frequency	51	76	57	49	30	93	356
	Row %	14.33	21.35	16.01	13.76	8.43	26.12	(34%)
	Column %	37.78	30.04	46.34	46.23	28.85	28.53	
Total	**Frequency**	135	253	123	106	104	326	1047
	%	12.89	24.16	11.75	10.12	9.93	31.14	100

Both the fact of numerous intersections of solution profiles in Table 1 and a formal model test of a Rasch scale (Andersen LR Test score = 104.99; df = 11, $p < 0.0001$) reject a homogenous latent trait as adequate psychometric model of PM comprehension, as measured by the given 12 items (see Rost, 1988, Andersen 1973). It is therefore not meaningful to interpret the sum of correctly solved items as a simple measure to quantify a latent, continuous ability of high-school students to understand graphical models. Instead, it seems necessary to compare the interrelations of the typical comprehension patterns as qualitatively differing groups according to other variables like sociocultural background and task-relevant eye movements.

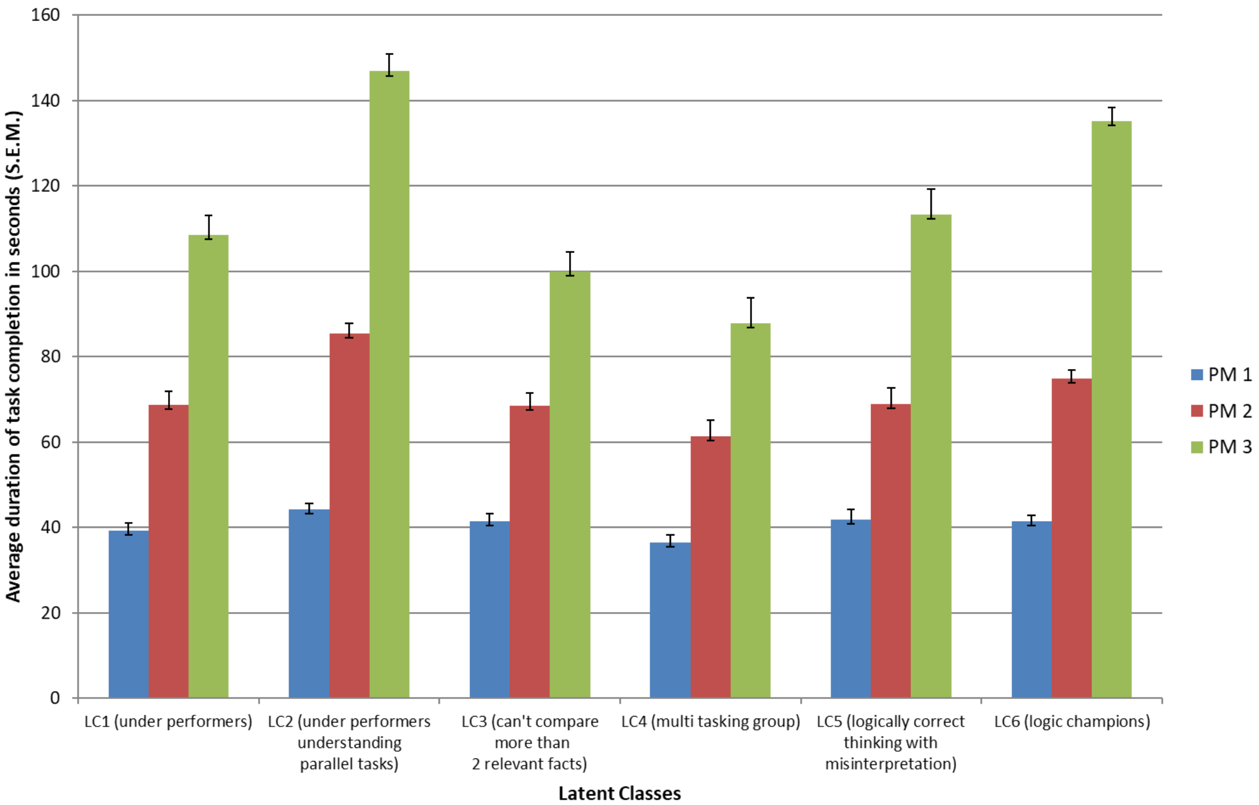
When events and decisions were presented under the “P”-condition (pseudo sentences), latent classes 3 (binary thinking group) and 4 (multi-tasking group) were more prevalent (each by 12%) than expected under the assumption of having no association between model condition and problem-solving pattern (see Table 2), while the better performing groups LC5 and LC6 were under-represented. Thus, describing processes with pseudo sentences seems to prohibit correct deciphering of more complex loop structures. When PM were presented with meaningful sentences (condition “S”), latent classes 2 (under performers with understanding of simultaneous tasks) and 5 (misinterpretation of parallel paths) were clearly over-frequented (by 15% and 26% respectively). Finally, under the condition of solely mentioning letters for events and decisions of a PM (condition “L”), latent class 6 (logic champions) was the most prominent cognitive solution pattern, and a clear under-representation of LC3 (binary thinking) and LC5 (misinterpretation of parallel paths) was observed. Denoting PMs with only letters thus favours good task performance. These effects are statistically significant (Pearson X^2 (d.f. 10) = 202.99; $p < 0.0001$) and can be interpreted causally, as each participant’s allocation to one of the conditions was randomly chosen.

Neither age nor gender of the participants, nor parental educational background or students’ self-ratings of being gifted with visual imagination could be shown to interact with class membership (results not shown here). The condition of PM presentation clearly resulted in differing durations of problem solving. Overall, task completion for the letters condition required, on average, 206.2 seconds (SD = 85.8) and meaningful sentences 239.2s (SD=82.0). In turn pseudo sentences required a mean duration of 290.7s (SD=149.9) before completely responding to all 12 items.

Increasing complexity of PMs required more time over all six latent classes ($F_{2,2080} = 2059.7$, $p < 0.001$). Though differences between latent classes ($F_{5,1040} = 16.3$, $p < 0.001$) and an interaction effect of complexity*latentclass ($F_{10,2080} = 30.8$, $p < 0.001$) in the respective ANOVA model proved also significant, this is mainly due to the large sample size. Effect sizes were 0.30 (eta squared) for complexity, but only 0.06 for latent classes and 0.03 for the interaction effect.

Figure 3

Impact of increasing complexity of PMs on task completion durations (=response latencies) in Sample I



Solution patterns and corresponding eye movement parameters in Sample II

Table 3 displays descriptive statistics for the eye-tracking measurements broken down by a) status of respondents' expertise, b) condition of the PM phrasing, and c) membership of the respondents in latent class.

Table 3. *Descriptive Statistics for the Eye-tracking Measurements in Sample II.*

		Expertise Status		Model Condition			Membership in Latent Class			
	Total sample II (N=36)	VL Experts (N=21)	VL Novices (N=15)	Letters (N= 14)	Sentences (N= 12)	Pseudo (N= 10)	LC4 (N= 6)	LC5 (N= 11)	LC6 (N= 16)	Other (N=3)
	Mean (*SD*)	Mean (*SD*)*[2]		Mean (*SD*)*[3]			Mean (*SD*)			
Response latency (sec)	78.10 (33.14)	**87.07 (30.66)**	**65.55 (33.37)**	**60.92 (26.49)**	**84.34 (36.61)**	**94.66 (28.36)**	57.80 (9.25)	85.40 (37.12)	82.12 (35.62)	70.52 (29.26)
Fixation duration on models (sec)	38.15 (19.91)	41.51 (17.87)	33.44 (19.91)	27.74 (15.40)	43.82 (23.99)	45.91 (14.76)	24.10 (5.05)	41.38 (19.45)	42.30 (22.84)	32.23 (15.22)
Fixation duration on models (%)	47.54 (7.39)	46.59 (7.52)	48.87 (7.39)	44.49 (7.28)	50.29 (8.14)	48.50 (5.42)	42.00 (7.23)	47.77 (4.89)	49.75 (7.21)	46.00 (13.45)
Fixation duration on Relevant (Red)*[1] (sec)	29.30 (17.86)	29.64 (13.1)	28.83 (23.48)	28.23 (15.02)	32.62 (25.65)	26.83 (9.31)	21.65 (8.67)	29.26 (12.41)	33.67 (23.38)	21.46 (10.14)
Fixation duration on Irrelevant (Blue)*[1] (sec)	10.41 (8.31)	12.14 (7.64)	7.98 (8.86)	**4.01 (2.37)**	**13.70 (9.53)**	**15.42 (6.59)**	5.00 (1.87)	9.96 (8.05)	12.29 (9.57)	12.86 (7.92)
Fixation duration on statements (sec)	25.81 (9.81)	**29.71 (9.68)**	**20.34 (9.82)**	21.29 (8.55)	25.79 (9.67)	32.16 (8.84)	22.91 (5.67)	28.13 (11.14)	25.29 (10.01)	25.82 (13.32)
Fixation duration on statements (%)	33.96 (6.45)	34.98 (6.73)	32.54 (6.45)	35.22 (3.80)	32.21 (9.83)	34.31 (4.03)	39.45 (6.05)	33.47 (3.11)	31.64 (6.70)	37.20 (10.09)
PM2 fixation duration on statements (sec)	21.37 (8.93)	24.44 (8.03)	17.06 (8.55)	18.99 (8.39)	22.43 (9.86)	23.42 (8.64)	20.32 (8.13)	20.90 (8.70)	22.46 (10.16)	19.59 (7.72)
PM2 fixation duration on statements (%)	29.76 (8.89)	31.09 (8.84)	27.90 (8.93)	30.26 (7.18)	29.32 (12.74)	29.57 (5.81)	35.14 (9.19)	28.75 (6.99)	27.99 (8.39)	32.10 (16.55)
Number of toggles	19.87 (8.24)	21.76 (9.34)	17.22 (8.24)	19.83 (9.34)	18.97 (9.41)	21.00 (5.24)	13.61 (3.55)	23.15 (9.63)	20.79 (7.06)	15.44 (10.36)
Rate of toggling	.267 (.083)	.253 (.083)	.287 (.092)	.333 (.078)	.223 (.067)	.228 (.040)	.239 (.061)	.295 (.102)	.271 (.074)	0.20 (0.07)

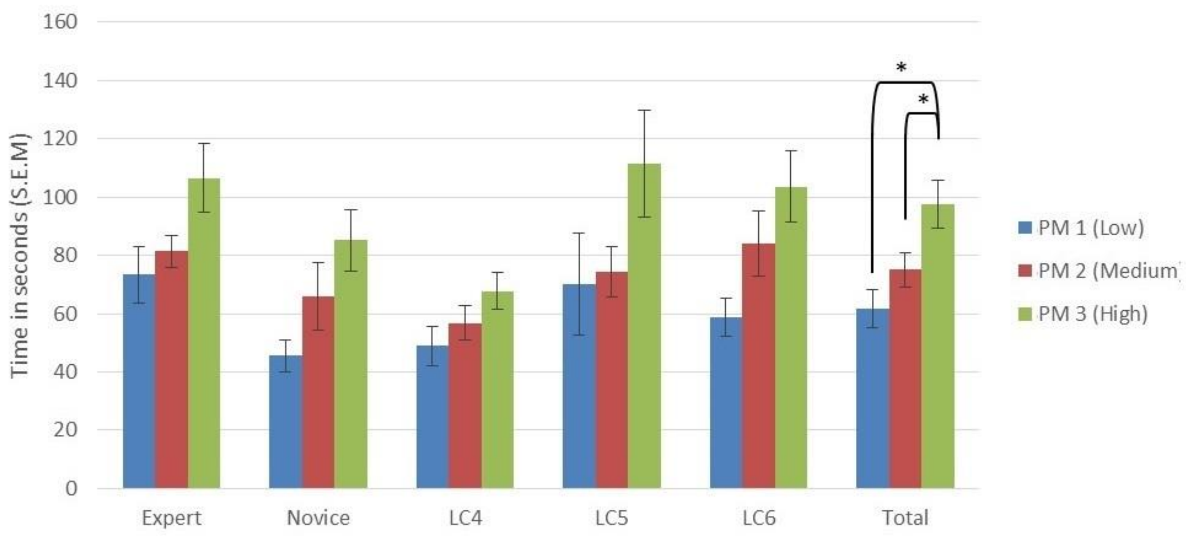
Note. *[1]exclusively for PM2, *[2] bold font: significant $p< .05$ for fixation duration on statements, marginally significant $p=0.053$ for response latency (t-test), *[3] bold font: significant $p< .05$ (F-test)

As in Sample I, increasing model complexity required longer response latencies ($F_{2, 70} = 12.31$, $p < 0.001$, $\eta^2 = .260$). With rising complexity, the fixation duration on models rose as well ($F_{2, 70} = 31.46$, $p < 0.001$, $\eta^2 = .466$) and the number of toggles increased ($F_{2, 70} = 7.49$, $p = .001$, $\eta^2 = .181$).

Bonferroni-adjusted post-hoc analysis revealed a significant difference in response latency between PM1 and PM3 (-36.44 sec., nominal p = 0.001) and between PM2 and PM3 (-22.11 sec., nominal p= .002) (see Figure 4). Additionally, number of toggles for PM3 was significantly higher than for PM1 (+ 7.6 toggles, nominal p = 0.004). Furthermore, response latency in the letter condition differed significantly from the one in the pseudo sentences condition (-33.74 sec., *p*< .05) with an average duration being about 34 seconds longer in the pseudo sentences compared to the letter conditions.

Figure 4

Response Latency in Seconds (SEM) on each Model by PM Complexity, Expertise Level, and Latent Class Membership in Sample II



Note. *significant (on *p*< .05)

No differences could be shown between VL experts and novices concerning eye movements, with the exception of fixation duration on statements, which differed significantly with VL experts spending more time on the possible responses than novices ($M_{Experts} = 29.71$ sec., $M_{Novices} = 20.34$ sec.; $F_{1,34} = 6.994$, $p < 0.05$, $\eta^2 = 0.171$). Also, task completion duration of VL experts tended to last longer (p = 0.053). VL experts tended to invest more time in arriving at any solution, but failed to outperform novices. There were no statistically significant differences between the VL experts and novices in fixation durations on relevant ($F_{1,34}$ = 0.017, n.s.) or redundant model parts ($F_{1,34}$ = 2.274, n.s.) of PM2,. We could also not demonstrate an association between expertise status and latent class membership (χ^2 (3, N = 36) = 1.870, *p* =.600). The number of toggles between PM2 and statements was inversely predictive for LC4 (OR = 0.785 [0.622-0.992]). Other eye-tracking measurements (fixation durations on either part of the model) were not associated with membership in latent classes. Membership in latent class, model condition, and visual expertise did not interact significantly with the main effect of increasing complexity. However statistical power is quite low for most of the variables in Table 3 (e.g. 1-β ranging from 0.069 up to 0.643 for the observed differences).

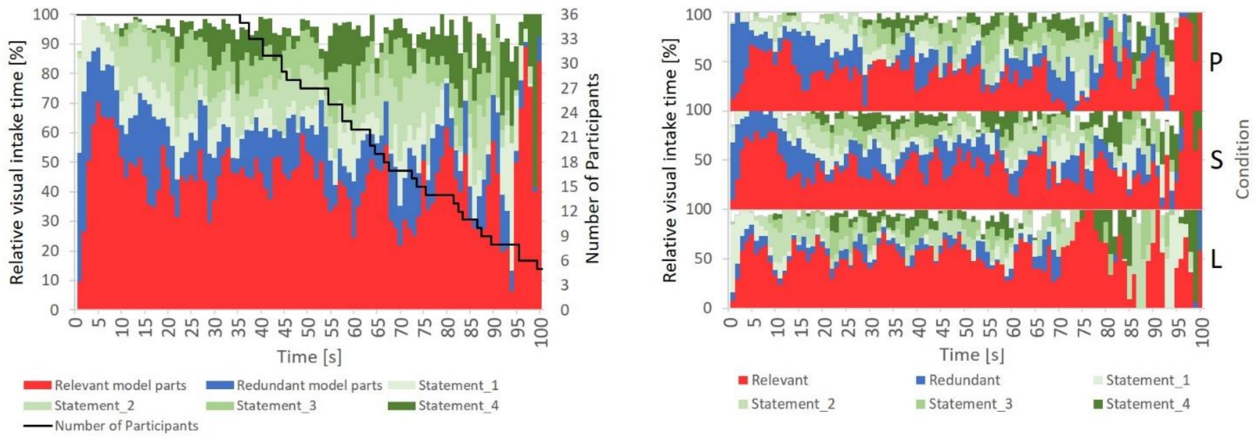
For a hypothetical “small” effect size in variable “response latencies (Cohen’s *d* = 0.22), meaning that experts were on average 7 seconds faster than non-experts, statistical power would reach 0.16. For a medium effect size (*d* = 0.40, 14 seconds difference) power would reach 0.35, and for a large effect size (*d*=0.66, 21 seconds difference) power would reach 0.60.

Fig 5A and 5B show the AOI hit distribution over the first 100 seconds of PM2. Different colours represent different AOI (see Fig. 2). As can be seen from Figure 5A, median response latency of PM2 (right vertical axis, solid black step function) in Sample II was reached in about 66 – 70 seconds. After this time, 50% of all participants in Sample II had made their decision for PM2, only 5 participants needed longer than 100 seconds to respond. PM2 was chosen as an example, as it proved to differentiate between the participants’ problem-solving patterns in Sample I most prominently. On average, participants directed their fixations primarily to relevant parts (red) of the model (29.3 sec.; SD 17.9), which is about three times longer than the time inspecting the irrelevant parts (blue) of PM2 (10.4 sec.; SD 8.3).

However, as can be seen in Figure 5B, there were characteristic differences between the three model conditions in attention distribution as measured by fixation durations.

Figure 5A (left) and 5B (right)

Histogram of AOI Hit Distribution for PM2 over the First 100 Seconds (A) and by Model Condition (L=Letter, S=Sentences, P=Pseudo Sentences) (B)

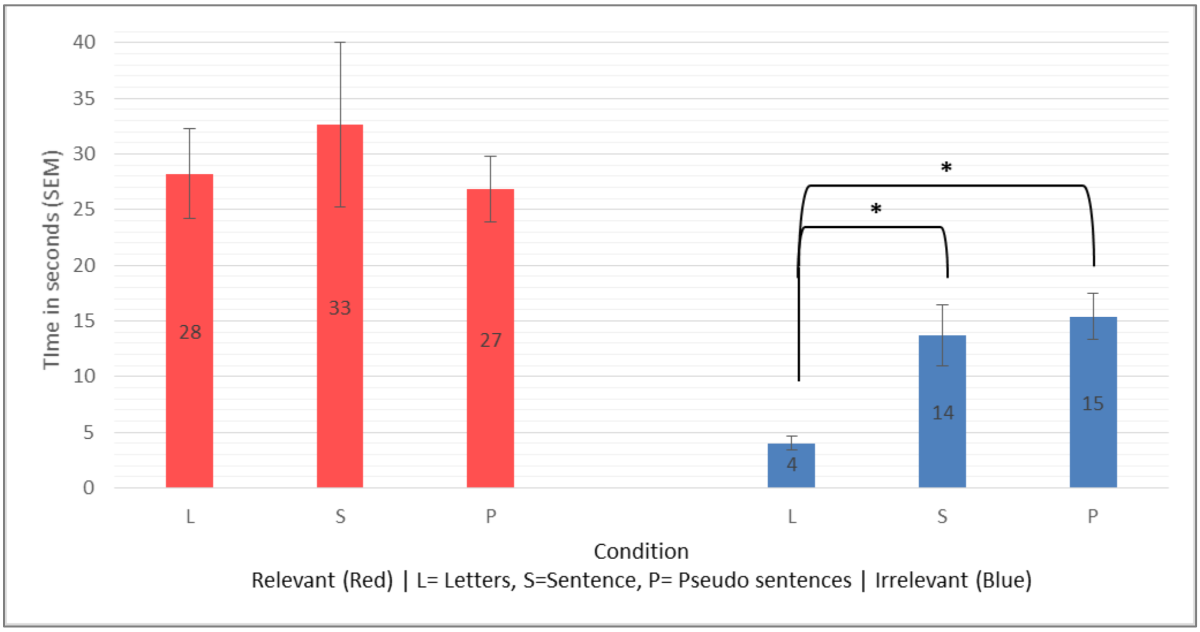


Further investigating the relationship between the different model conditions (L, S, P) and the time spent on fixating different (relevant/irrelevant) parts of the PMs revealed an advantage of the letter condition with respect to the redundant parts of the model: Separately analysing fixation durations by model condition (Figure 6) indicates that the letter condition is associated with shorter fixation periods on irrelevant parts of the process model (*M*= 4.01 sec., *SD*=2.37) compared to the sentence (*M*=13.70 sec, *SD*= 9.64) and pseudo sentence (*M*=15.42 sec., *SD*=6.59) condition ($F_{2, 33}$ =10.757 sec., $p<.05$, η^2= 0.395).

Figure 7 illustrates the total time spent on the process model (= response latency, left half A) and fixation duration on each process model (right half B) as part of the total response latency.

Figure 6

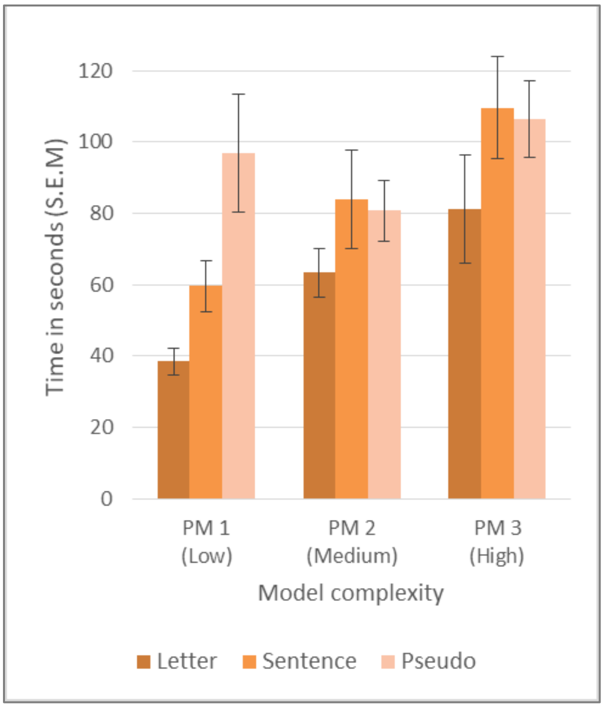
Average Fixation Duration on Relevant and Irrelevant Parts of PM2 by Condition

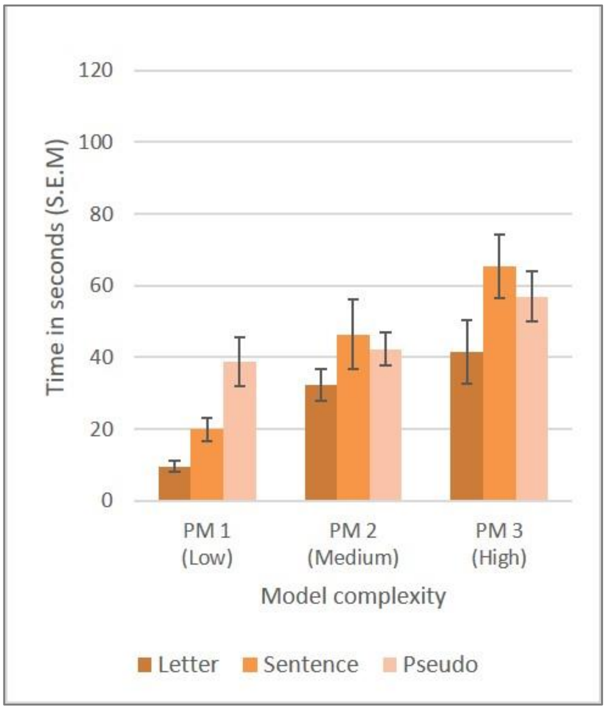


Note. *significant (on *p*< .05)

Fig. 7A (left) and 7B (right)

Bar Charts of Average Response Latencies (left) and Average Fixation Duration (right) on each PM by Model Complexity and Letter, Sentence and Pseudo Sentence Condition





4 Discussion

Measurement of PM comprehension: Solution Patterns

Six latent classes with qualitatively differing solution profiles were adequate to classify scholars in Sample I. These configurative and non-ordered profiles can be interpreted as separate solution patterns, where specific model parts are understood better than others. Beyond very good performers (LC6 “logic champions”) and quite bad performers (LC1 “under performers”) there exist other groups of students at intermediate “levels”, which can be related to qualitatively differing errors. E.g. isolated good comprehension of simultaneous activities in process models (LC2) in front of otherwise bad performance, or isolated lacking comprehension of parallel paths (LC5), or lacking capacity to compare more than 2 relevant facts (LC3). Participants in LC4 are best in understanding the concept of parallel pathways, but at the same time do not easily understand repeating loops.

Thus, an interpretation of the total number of correct responses would disregard important differences between different cognitive strategies mainly for “average” good participants. Given the unknown increase in cognitive workload with more complex graphical models, and given the experimentally varied wording conditions of graphs and test items, and thirdly given the differing logical problems formulated by test items, a grouping algorithm like LCA seems to be a good choice to differentiate students according to their capacity to decipher process models.

Moreover, differentiating specific comprehension errors has also a practical implication: Within educative context, it is important to know, which specific concepts and tasks are still misunderstood or are already understood in order to give meaningful feedback (Shute, 2008). Knowing which solution profile a learner applies helps to give meaningful feedback and derive adequate strategies for improvement.

In Sample II, the majority of the participants responded in a similar fashion to the profiles of LC5 (“logically correct thinking, with misinterpretation”) or LC6 (“logic champions”). This better performance might partly be explained by the higher mean age and, resulting from that, the longer formal education of these participants. Nevertheless, solution patterns were only weakly connected to aspects of eye movements while working on the tasks. Only the number of toggles (gaze transitions) between the graphical model and the written statements was negatively associated with membership in LC4 (“multi-taskers”). The lower the number of toggles in PM2, the more likely the participant displayed a correct understanding of parallel pathways (even better than LC6), while failing to understand the notion of loops. In other studies a high rate of toggling was negatively correlated with intelligence scores that used visual tasks as a measurement basis (e.g. the Wiener Matrizen Test 2, see (Laurence et al., 2018)). Excessive toggling characterized a strategy to eliminate mutual contradictory responses instead of finding logical sequences within systematically ordered matrices of pictograms (Arendasy & Sommer, 2013; Bethell-Fox et al., 1984). In our study, the four statements underneath each PM often addressed similar activities. In PM2 there were two statements addressing the notion of loops, which could have been weighted against each other by means of toggling (Q1: "E must be executed at least once" vs. Q3: "E can be executed a maximum of four times").

Even though LC5 and LC6 were quite different in the comprehension of PM2, other eye-tracking measurements like the participants' fixation durations on either part of the model (classified into various areas of interest) were not associated with membership in latent classes. But finding no differences could be due to low statistical power.

Features impacting comprehension: PM complexity

Model complexity was handled as a within-subject factor in each condition. With increasing model complexity, the time required to respond to the comprehension questions rose. This is true for both Sample I and Sample II. Concerning eye movement indicators, the same increase could be observed for fixation duration on the models and the total number of toggles. This demonstrates that participants aspired to find the correct solutions and were not prone to click a response alternative quickly or randomly, in reaction to overly excessive demands. While we do not have comparable eye-tracking data in Sample I, the participants had been asked whether they thought the test was too difficult to be solved and whether they understood the tasks. Only 25 participants (of 1047) responded in the affirmative to the former and 23 denied the latter question. Therefore, we assume a high aspiration level across both samples, which supports a preliminary interpretation of the determined latent classes as potential "cognitive styles". It should be kept in mind, that the interpretation of latent classes as "cognitive styles" is based on a purely data driven approach and should be regarded a preliminary tentative interpretation of empirical solution patterns. Further studies should focus on a convincing link between cognitive theory and solution patterns, as the latter might change with alternative operationalisations of PM complexity.

Features impacting comprehension: Semantic notation

The PM conditions, i.e., whether the PM components had been labelled by letters, sentences, or pseudo sentences, were associated with a different prevalence of latent classes in Sample I (see Table 2). They also exerted a systematic influence on some of the eye-tracking variables. Contrary to our expectations, sentences representing everyday processes as naturalistic scenarios were not associated with a higher prevalence of the “logic champions” LC6, as earlier studies would have predicted (Sweller & Sweller, 2006; van Merrienboer & Sweller, 2005). Instead, in more than half of the participants single letters as denotation generated a solution pattern of the “logic champion” type. This is in line with the finding of Mendling et al. (2012) on the impeding effects of additional semantic information on syntax comprehension.

Stimulus features nested in the PMs appear to impose a high extraneous cognitive load (Sweller, 2005) that requires working memory resources. Longer fixation duration (as measured in Sample II) can be understood as prolonged cognitive processing (Sweller et al., 2011, p. 81). eye-tracking data can indicate where and for how long the subject focuses his or her attention, implying corresponding variations on cognitive load. When splitting the model AOI into relevant and irrelevant parts, as we did with PM2 (see Fig. 2), the fixation duration on irrelevant parts was significantly shorter in the letter condition than in the sentences and pseudo sentences condition. On the other hand, fixating relevant parts of the models displayed no significant differences between conditions (see Fig. 6). The relevant parts all had about the same fixation time in all three stimulus conditions (see the percentage of red and blue in Fig. 5B).

Additional verbal workload, regardless of sentences content (pseudo or real sentences) does not increase the time needed to focus on relevant model parts; additional time is only spent on verifying irrelevant model activities. Verbal attributes seem to distract from identifying the relevant model parts, but do not increase the time needed to focus on the relevant parts of the model. One might assume that for PMs that only include letters, the fixation duration could be expected to decrease on every part of the model corresponding to less reading time. However, this is not the case here. So, what contributes to this effect?

We assume three different types of cognitive processes, which are needed to come up with a solution to the statements presented below each model. First you need to *read and understand* (A) the sentences and model activities, then *find and compare* (B) the statements with the relevant model parts, and finally *evaluate and decide* (C) whether or not the statement is correct. This follows the idea of the so-called SOI model (“Selection-Organization-Integration”), which has been elaborated for cognitive load theory in multimedia learning (Mayer, 1996, 1999). The time spent on irrelevant parts is only used for *reading and understanding* (A) as well as *finding and comparing* (B), but not for *evaluating* (C) the statements. A and B take significantly longer in the sentences and pseudo sentences condition as the structure of the sentence and the meaning of words need to be understood before it can be rejected as irrelevant. The relevant parts of PM2 include logical gateways, which were essential for answering most questions. These gateway symbols did not differ between conditions. From this point of view the fixation duration on relevant model parts should not differ between conditions, as the symbols did not change between conditions and the time spent on relevant model parts prominently included the time to *evaluate and decide* (C), whether the statement is true or false.

It might be speculated that a model, which combines letters for redundant model parts and sentences for important model parts, would be the most efficient design implementation for reducing the time spent fixating on the model as a whole. The practical implication would be that the most important information can be presented in a more natural verbal form (sentences), where other information should be presented in a short “logic-inducing” variant (e.g. letters or symbols) to keep the observant from looking at less important model parts and therefore reducing cognitive workload of *reading and understanding* (A) as well as *finding and comparing* (B). Further research needs to be conducted, in combining both elements in one process model to verify these conclusions.

Visual Literacy and PM comprehension

We could not find significant differences in cognitive solution patterns between VL experts and novices. Thus, understanding and “solving” process models does not seem to depend too much on visual literacy as defined in this study. Apparently, comprehending the logic behind IF and OR gates as well as recognising pathways is crucial to follow the information flow in PMs. Even though the PMs are presented in a visual form, the ability to “interpret, analyse or appreciate visual media” does not seem to help understanding the “logical structure” of the PM. This result is useful with respect to other VL assessment items in terms of discriminatory validity. Given the small observed mean differences, it seems reasonable to hypothesize that the capacity for solving PMs does not contribute to the distinctiveness of visual literacy, which brings up an important distinction between logical models and other forms of visual information (e.g. parts/details of pictures Vogt & Magnussen, 2007). Regarding the eye-tracking indicators, we also did not find significant differences between VL experts and novices on fixation duration between relevant and irrelevant PM parts. If VL had a substantial influence on PM comprehension, we would assume longer fixations on relevant AOIs and shorter on irrelevant AOIs, as indicated by Gegenfurtner et al. (2011). On the contrary, it seems that the search for subjective factors impacting PM comprehension (favoured by Recker and Dreiling, 2011) should not address primarily visual competence but cognitive capacities.

VL experts spent more time looking at the four statements below each model, and therefore took more time reading or thinking about the given statements. It would be interesting to see if artistic model features like colours or fonts would facilitate or distract specifically VL experts in following the logical character of PM. In further studies, longer linear models (requiring the exclusion of more nodes as “irrelevant”) could help to distinguish between the workload emerging from actively omitting irrelevant facts from the workload necessary to draw logical decisions. That way the effect of verbal contribution on the distribution of cognitive load could be differentiated independently from the influence of logical gateway symbols.

Practical applications and future investigations

Eye tracking allows for a multitude of interesting experiments on analysing visual perception (Holmqvist & Andersson, 2017). Many other studies try to find differences in eye-movements between experts and novices. Experts in their field may faster distinguish relevant from irrelevant information than novices do (Gegenfurtner et al., 2011). For example, it can be shown that expert chess players are able to use their parafoveal vision (complete field of vision) to extract information that is relevant for the solution of the tasks better than novice players (Charness et al., 2001; Reingold et al., 2001; Sheridan & Reingold, 2014). Higher cognitive functions like this holistic perception of a scene require perceptive as well as memory processes. Whether or not the VL experts in our study profit from their greater experience with visual stimuli or whether they were able to perceive relevant details more holistically, should not be decided on our novel setting, because the perceptive part of the visual tasks may be mantled by necessary logical reasoning.

There are implications that could lead to practical progress e.g. in teaching software engineering. The video recordings of participants gaze behaviour on target stimuli can be used as an educational tool, to show and teach novices when and where to look at (e.g. in information retrieval from medical images; Gegenfurtner et al., 2017). Combining eye-movement modeling examples (known as EMME) with other learning systems used for training in process model comprehension (e.g., a step-by-step assistant that teaches a complete and correct comprehension of process models) can be developed accordingly, thus enabling especially novices a better initiation to working with process models (see Jarodzka et al., 2017 for further proposals in using eye tracking in educational context).

The identification of latent classes with differing solution profiles helps to provide learners with useful feedback on adequate strategies how to improve their decisions. Assessment of visual competence might be helpful to address different target groups among apprentices while preparing specific learning materials (Andrà et al., 2009). We encourage further research on process model comprehension by means of eye tracking. Moreover, in the context of Industry 4.0, process models serve as an enabler for automatization. Because process models used for this purpose often are very complex and thus hard to read and comprehend, the methodology introduced by this article might contribute to enable further studies with high relevance for the field of organizational research (Meißner & Oll, 2017).

Limiting factors

Some limiting factors of our study need to be addressed. (1) We assume the same latent classification from Sample I (high-school students) to be present in Sample II (VL expert and novice group). However, it is possible that through age differences and recruitment outside a classroom context, different underlying classifications might be more appropriate. (2) When looking at AOIs from a narrow and dynamic visual angle, the risk for error prone AOI-fixation detection increases (Orquin & Holmqvist, 2018). Our AOIs were therefore drawn more conservatively (larger) and included multiple activities and pathways to compensate for eye-tracking inaccuracy. Using remote devices with constant lightning conditions and steady head position (minimizing Pupil foreshortening effect, Hayes & Petrov, 2016) in future studies could avoid this imprecision and also allow pupillometric analyses. (3) The generalizability of the typology of cognitive solution patterns to other PMs is difficult, if not impossible, due to the different features of the PM that we used to operationalize complexity. Increasing model complexity was based on the guidelines from Becker et al. (2000): PM1 was constructed as a linear model, PM2 had one prominent parallel pathway and one loop, and PM3 had multiple inclusive and exclusive pathways in combination with a higher number of total activities. Whether this selection of demand characteristics is representative for the whole universe of possible model complexities cannot be decided from our data.

(4) Potential effects of various statistical aspects of our study: sequence of model presentation and/or of comprehension test items cannot be excluded due to their uniform ordering

corresponding to their complexity. (5) The selection of valid eye-tracking indicators: At this point, we could not deduce a single variable as major study endpoint because of lacking theoretical foundation, and also could not construct a combined scaled measure of the correlated variables in use due to the limited sample size of study II. (6) The semantic language structure of the four statements presented below each PM was also not varied systematically: PM1 only included questions regarding sequence (e.g. A follows B), PM2 included questions regarding sequence, conditional activities and loops, and PM3 included questions on sequence, on conditional activities as well as a statement on all activities in the model (PM3, Q3). Therefore, it is difficult to identify a specific model feature or statement as exerting the main influence on the solution patterns. Future studies should systematically vary the cognitive workload that results from the logical structure, labels or comprehension statements.

(7) Finally, if done in more detail, the latent structure of solution patterns could be analysed using more sophisticated psychometric models than a “simple” latent class analysis. Though LCA seems appropriate for the comparisons in this study, it is conceivable that different subgroups of high-school students (or adults) share different PM features for comprehension. Mixed Rasch Models or so called “hybrid models” (Rost & Langeheine, 1997) may be applied to test these patterns of responses in PM comprehension tasks. As in research on intelligence, one could also speculate on the existence of second order abilities (dominated from the subject’s characteristics) and first order task-specific latent classes.

Conclusion

To conclude, the present study demonstrates an association between problem solving behaviour as measured through eye tracking and the comprehension of PMs. Specific solution patterns could be revealed, depending on the structure and complexity of PMs. The condition of how PMs are presented (i.e., letter, sentence, or pseudo-sentences) displayed significant influence on the answering patterns and the time spent on each model. PMs cannot be interpreted solely based on their graphical nature, but their semantic structure plays an important role for their comprehension as well. Specifically, the use of single letters for model activities resulted in a faster and more precise understanding of the models. Experts in VL could not be shown to outperform novices with respect to PM comprehension. It seems worthwhile to focus on the cognitive mechanisms and less on visual competence of subjects when assessing their PM comprehension.

From a methodological point of view, eye tracking demonstrated a fruitful path into analysing the comprehension of graphical logical models like PMs. Fixation duration on different parts of a model enabled scrutinizing effects of verbal model features on attention distribution and cognitive workload. In future studies, relevant and/or difficult to comprehend parts in a process model may be extended with other visual features for effective guidance through a PM. Due to the restricted variation of characteristics of (Business) PMs, further research needs to include a wider range of model formulations.

5 Author contributions

UF and KR designed the study. MW, MR, and RP designed the process models, their presenting conditions, and constructed the related answering statements testing comprehension. MW and MT wrote the program, under supervision of MR and RP, and implemented it on the tablets. MT performed the field work. MT and MG designed and analysed all eye-tracking measurements. MT and UF performed all psychometric and other statistical analyses. MT, UF and KR prepared the manuscript. All authors read the manuscript and reviewed it critically. All authors listed thus have made a substantial, direct and intellectual contribution to the work, and approved it for publication. We would also like to thank Kenneth Holmqvist for his helpful comments on an earlier version of this paper.

6 Funding

This work was supported by the German Ministry of Education and Science (grant number: 01JK1606A). The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be interpreted as a potential conflict of interest.

Paper 2: Visual Search on Artworks

How do Art Skills Influence Visual Search? – Eye Movements Analyzed with Hidden Markov Models

Miles Tallon, Mark W. Greenlee, Ernst Wagner, Katrin Rakoczy, Ulrich Frick

Author-produced version of this article published after peer-review on January 28th 2021 in Frontiers in Psychology: https://doi.org/10.3389/fpsyg.2021.594248 (Tallon et al., 2021) reprinted with permission.

Abstract

The results of two experiments are analyzed to find out how artistic expertise influences visual search. Experiment I comprised survey data of 1065 students on self-reported visual memory skills and their ability to find three targets in four images of artwork. Experiment II comprised eye movement data of 50 Visual Literacy (VL) experts and non-experts whose eye movements during visual search were analyzed for nine images of artwork as an external validation of the assessment tasks performed in Sample I. No time constraint was set for completion of the visual search task.

A latent profile analysis revealed four typical solution patterns for the students in Sample I, including a mainstream group, a group that completes easy images fast and difficult images slowly, a fast and erroneous group, and a slow working student group, depending on task completion time and on the probability of finding all three targets. Eidetic memory, performance in art education and visual imagination as self-reported visual skills have significant impact on latent class membership probability. We present a hidden Markov model (HMM) approach to uncover underlying regions of attraction that result from visual search eye-movement behavior in Experiment II. VL experts and non-experts did not significantly differ in task time and number of targets found but they did differ in their visual search process: compared to non-experts, experts showed greater precision in fixating specific prime and target regions, assessed through hidden state fixation overlap.

Exploratory analysis of HMMs revealed differences between experts and non-experts in image locations of attraction (HMM states). Experts seem to focus their attention on smaller image parts whereas non-experts used wider parts of the image during their search. Differences between experts and non-experts depend on the relative saliency of targets embedded in images. HMMs can determine the effect of expertise on exploratory eye movements executed during visual search tasks. Further research on HMMs and art expertise is required to confirm exploratory results.

Keywords: visual literacy, assessment, fixation sequence, hidden markov model, eye-tracking data, visual search task, latent profile analysis

1 Introduction

Visual perception is an active process of constructing meaningful information from external visual stimuli based both on neurobiological capacities (i.e. laws of perception) and individual learning history (skill training, memory). Perceptual psychology describes the cognitive mechanisms employed to transform visual stimuli into information. The comparison of experts' and non-experts' processing during a challenging visual task can be used to decipher these cognitive mechanisms. In a broad sense, visual expertise has been studied in medicine (medical imaging), engineering (surveillance of technical processes) or education (learning behavior) and has been defined as a domain-specific adaptation to the requirements of a visually challenging task (Gegenfurtner & van Merriënboer, 2017), which has been coined Visual Literacy (VL). More recently, mostly from authors in the context of aesthetics and fine arts, this concept has also been referred to as visual competency (Schönau & Kárpáti, 2019). Other authors (Avgerinou & Pettersson, 2011; Wagner & Schönau, 2016) used the term VL, which they described as the ability to inspect and understand images and express oneself through visual media.

Psychological models of visual expertise have focused on three major theories (Brams et al., 2019; Gegenfurtner et al., 2011): (1) the long term working memory theory (Ericsson & Kintsch, 1995) suggests that experts can retrieve more visual information from long term working memory than novices do, (2) the information reduction hypothesis (Haider & Frensch, 1996, 1999) proposes that experts selectively focus on important visual image parts relevant for the task and ignore irrelevant stimuli, and (3) the holistic model of image perception (Kundel et al., 2007), which states that experts gain more visual information from global and para-foveal regions, effectively allowing a broader grasp of the image to guide their search. Recent studies find evidence in support of the information-reduction hypothesis as the most important skill developed in experts across most domains (Brams et al., 2019).

Restricting the discussion to fine art studies and art education, differing VL models have been proposed (Kędra, 2018). One of the broadest conceptual models (ENViL-model, see Wagner & Schönau, 2016) divides VL into as many as 16 subdomains, which include "value", "envision", "experiment" or "aesthetic experience". Many of these domains show considerable overlap. As this model was not generated by psychometricians but by art educators, this model is strictly phenomenological and has not been empirically tested. Nevertheless, it has received great attention from applied art education theory (e.g. Groenendijk et al., 2018). One subdomain, "analysing", has been described as the ability to attentively and accurately focus on visual stimuli and to identify characteristics of images (Wagner & Schönau, 2016, p.70) and therefore should be crucial in visual search tasks. The "analysing" ability is directly associated with the information-reduction hypotheses, where experts find important visual information by focusing on important image features and ignoring irrelevant features. Experts' continuous engagement in art and imagery plausibly should impact on how cognitive processes differ between VL experts and non-experts. VL experts thus may serve as best-practice examples to describe effective cognitive strategies in detecting details in images of artwork.

In a visual search experiment the participant is asked to look for a target among distractors (Wolfe et al., 2003; Wolfe, 2010). A visual search experiment might be analyzed by either evaluating the correct solution or by recording task-solving behavior (e.g., reaction times, eye movements). Thus, visual search paradigms are oftentimes used to investigate differences of experts and novices with respect to the participant's speed or accuracy in locating targets, for example in medical image examination (Drew et al., 2013; Sheridan & Reingold, 2017; van der Gijp et al., 2017) or in sports (Piras et al., 2014; Vaeyens et al., 2007). The influence of *reading* literacy on visual search has been extensively studied (Ferretti et al., 2008; Franceschini et al., 2012; Olivers et al., 2013). Few studies have considered the influence of *visual* literacy and its effects on visual-search performance. Studies on artistic visual expertise (e.g. Francuz et al., 2018; Vogt & Magnussen, 2007) are typically not conducted with visual search tasks (finding targets among distractors) but e.g. by judging abstract from realistic paintings or in the context of visual memory tasks. Expertise-related differences in target search have been mainly explored in domains other than the visual arts (e.g. medicine (Kundel et al., 2007) or sports (Vaeyens et al., 2007)). To our knowledge, the use of artwork in a visual search task still remains fairly uncommon (e.g. Nodine et al., 1979).

Aesthetic appreciation and a general interest in the visual arts might amplify a person's ability to identify specific details in images of artwork. With respect to art appreciation, five domains have been put forward: A (attraction), R (representation and realism), E (emotional expression) S (style and form), and I (interpretation), denoted as ARESI classification (van Meel-Jansen, 2006). Visual experts tend to show more appreciation for images rated high on the "style and form" and "realism" domain. Evidence from neuroaesthetic research revealed perceptual processing enhancement at behavioral and at a neurophysiological level when images are aesthetically appreciated (Sarasso et al., 2020). Art appreciation might facilitate visual search performance. Research on the aesthetic appreciation of art has differentiated between two modes of perception: pleasure and interest, which are conceptualized as partly overlapping, partly distinct functions of aesthetic appreciation (Graf & Landwehr, 2017). Whereas the free viewing of abstract paintings presumably favors the “pleasure” mode, a visual search task for a specific detail in paintings might require a more analytical way of regarding a piece of art (“interest” mode), depending on the painting style or content of the artwork. This might also elicit aesthetic appreciation. Thus, experts would operate more according to the “interest” pathway to aesthetic appreciation.

Studies of eye movement behavior often are regarded as a tool to link observed behavior to cognitive mechanisms (Hollingworth & Bahle, 2020). Previous eye-movement research has focused on domain-specific differences in visual expertise with respect to number of fixations and fixation duration (Gegenfurtner et al., 2011). Studies showed that professional art viewers were reported to exhibit greater saccadic amplitudes than novices, particularly when viewing abstract paintings (Zangemeister, 1995). Experts also tend to have more short fixation durations, i.e. direct their attention on specific areas of paintings (Ylitalo et al., 2016). Novices, when revisiting previously seen images, exhibit fewer and longer fixations (Vogt & Magnussen, 2007). Some studies used eye movement patterns to distinguish artists from laymen (Kolodziej et al., 2018). Other studies demonstrated differences between expert and novice artists in how they looked at particular works of art. Accordingly, experts differ from novices in the number of fixations and average fixation duration on specific parts of the image (Kolodziej et al., 2018). Interestingly, in a study on artists’ free viewing behavior of abstract paintings, Koide et al. (2015) report that expertise leads to fewer fixations on salient image regions. The authors suggest that the artists’ knowledge of art overrides stimulus-driven guidance of fixations, opening up the possibility of focusing their attention to less obvious image areas. Even though the transfer of VL ability across the art domain boundary remains uncertain, some studies have found differences in visual-spatial tasks depending on the person’s level of artistic expertise (Angelone et al., 2016; Chamberlain et al., 2019). In these studies, visual artists outperform novices through top-down control over attentional processes and fast and more precise visual encoding.

As previous studies pointed out differences in fixation duration or number of fixations, the time dependence of fixation sequences is rarely taken into account. However, the order of fixation sequences can be used to deepen our understanding of expertise-related differences in visual search. Eye-movements play an important role in visual search behavior, as they can indicate where and for how long people look at something, allowing researchers to model attention throughout the given task (Hollingworth & Bahle, 2020; Koochaki & Najafizadeh, 2018).

Heatmaps (Bojko, 2009), for example, can be used to visualize fixation density (number of fixations) over time. However, sequences (spatial as well as chronological order) of eye movements (i.e. scanpaths) are often neglected in the analysis of saliency or fixation density (Le Meur & Baccino, 2013). The sequence of eye movements is crucial in understanding not only where, but in what order people direct their gaze and attention while inspecting images.

To account for sequence dependent effects in eye movements, we chose to compare the spatial coordinates of fixations (i.e., the series of numbered fixations by index) across VL expert and non-expert groups. We use Hidden Markov Models (HMM (Rabiner, 1989)) to analyze the fixation sequence to reveal latent image areas. Through HMM-analysis we can find and visualize hidden (i.e. not directly observable) attention states (for details see methods section). Combining eye-tracking analysis with statistical approaches such as HMM leads to further insight into factors underlying scanpaths during visual search (Borji & Itti, 2013; Coutrot et al., 2018; Koochaki & Najafizadeh, 2018; Ulutas et al., 2019). HMMs have been successfully used in combination with eye-tracking data to parse fixations from saccades (Houpt et al., 2018), to depict processes underlying facial recognition (Chuk et al., 2014; Chuk et al., 2019) or for information retrieval during reading (Simola et al., 2008). The feasibility of a latent state approach for the analysis of eye-tracking data has become increasingly popular in applied research areas such as marketing research (Netzer et al., 2017).

Which parts of the image are closely examined and in what order are they examined during visual search? Search effectiveness and the probability of finding pre-defined targets are only one aspect of visual search performance. The psychometric assessment of search time and number of correctly identified targets does not allow for a detailed understanding of the underlying search process. Using eye-tracking measurements search processes and differences in expertise strategies can be examined more closely.

The research reported here is part of a larger study (Rakoczy et al., 2019) on visual literacy. A test battery was constructed to assess various aspects of VL, which was administered to a large sample of high school students for psychometric evaluation. The reliable and valid measurement of VL could serve as a tool for quality management in educational settings and thus contribute to the improvement of art education. Two aspects are especially interesting for the study of VL: First, not much is known about self-reported artistic skills and VL performance in young students. What influence do self-reported visual skills have on search time and number of found targets of students? Self-confidence or interest in visual arts may facilitate engagement in artistic stimuli. Secondly, cognitive mechanisms employed to solve visually guided tasks are a necessary link to translate skill level measurement to didactic improvements and to sharpen the association between self-perceived visual competency and art teacher's feedback. In addition to the psychometric evaluation of visually guided tasks, we compare VL experts and non-experts' visual search process to uncover expertise-specific modes of search behavior. Eye tracking is used to determine the external validity of the assessment items. Do VL experts differ from non-experts in search time and number of found targets? The comparison of both a student sample and a sample of VL-experts and novices can enhance our understanding of cognitive processes engaged in the visual tasks going beyond the measurement of performance (i.e., reaction times and hit rates; effectiveness) to also include information on order and precision of the search (efficiency).

This study addresses the following hypotheses:

H1a: VL experts identify more targets than students (Experiment 1)
H1b: VL experts are faster than students in finding the targets (Experiment 1)

H2a: VL experts identify more targets than non-experts (Experiment 2)
H2b: VL experts are faster than non-experts in finding the targets (Experiment 2)

To get insight into the search process we take a closer look at the participant's eye-movements during the search: Do VL experts differ in spatial and/or chronological aspects of their scanpaths from non-experts? More specifically, do VL experts identify more and/or other meaningful regions of interest in images of artwork than non-experts do?

H3: VL experts show higher precision in target detection, i.e. exhibit eye movements to targets that differ from those of non-experts during visual search (Experiment 2)

The eye-tracking research questions are assessed through exploratory analysis with the help of HMM models to investigate differences between the search strategies of VL experts and novices. Differences can be interpreted as empirically derived hypotheses for future confirmatory analysis. As the use of HMM in the context of expertise research is relatively new, we give some examples of how this method can be advantageous over traditional eye-movement visualization and analysis.

2 Method

Subjects

The data reported in this study was acquired as part of a larger research project on the assessment of Visual Literacy (Rakoczy et al., 2019) and is comprised of two samples: an assessment sample (Sample I) involving a large sample of high-school students and an eye-tracking sample (Sample II) consisting of VL experts and non-experts.

Sample I comprised 1065 high-school students from 52 classes (9th to 13th grade) of 29 schools in Germany of which 1056 worked on the visual search task. Overall, 52% were female, the average age was 15.27 years (SD = 0.94). Schools were recruited in the federal states of Hessen, North-Rhine Westphalia, Schleswig-Holstein, and Rhineland Palatinate via leaflets, letters and personal recommendations. The test was conducted in regular classrooms. Up to 30 students were able to participate in the assessment simultaneously. The visual search task under investigation was one segment of a longer (duration: 45 min) study on the topic of VL including a sociodemographic questionnaire (age, gender) and questions regarding the topic of art and personal experience with art: “Do you regularly attend an art school or art workshops?” Scale from 1 (never) to 4 (multiple times a week), “Art is important for me personally”, “My parents are interested in art and artistic subjects”, “In our family, art is very important”, “We like to talk about art and artistic subjects in our family” on a scale from 1 (strongly disagree) to 4 (strongly agree); “How good are you at art theory (e.g. interpreting pictures, understanding art history)?” “How well do you perform in arts education generally?” from 1 (very bad) to 5 (very good) and questions including the grade in art class and self-reported skills: photographic memory (PM; “I have a 'photographic memory'”), spatial orientation (SO; “When I see a photographed geometric object, I can imagine what it looks like from behind”), long-term memory (LM; “I can remember small details in pictures”), imagination (IM; “I can visualize things mentally”), and interest in visual puzzles (IP; “I like to solve picture puzzles”). These were reported on a scale from 1 (strongly disagree) to 4 (strongly agree). All answers were given via touchscreen input by the participants. School classes were offered a lump sum of 100€ as collective compensation. Sample I was presented four images included in the visual search task: “Exhibition”, “Oppermann”, “Footprints” and “Clock & Graffiti” (see Figure 1).

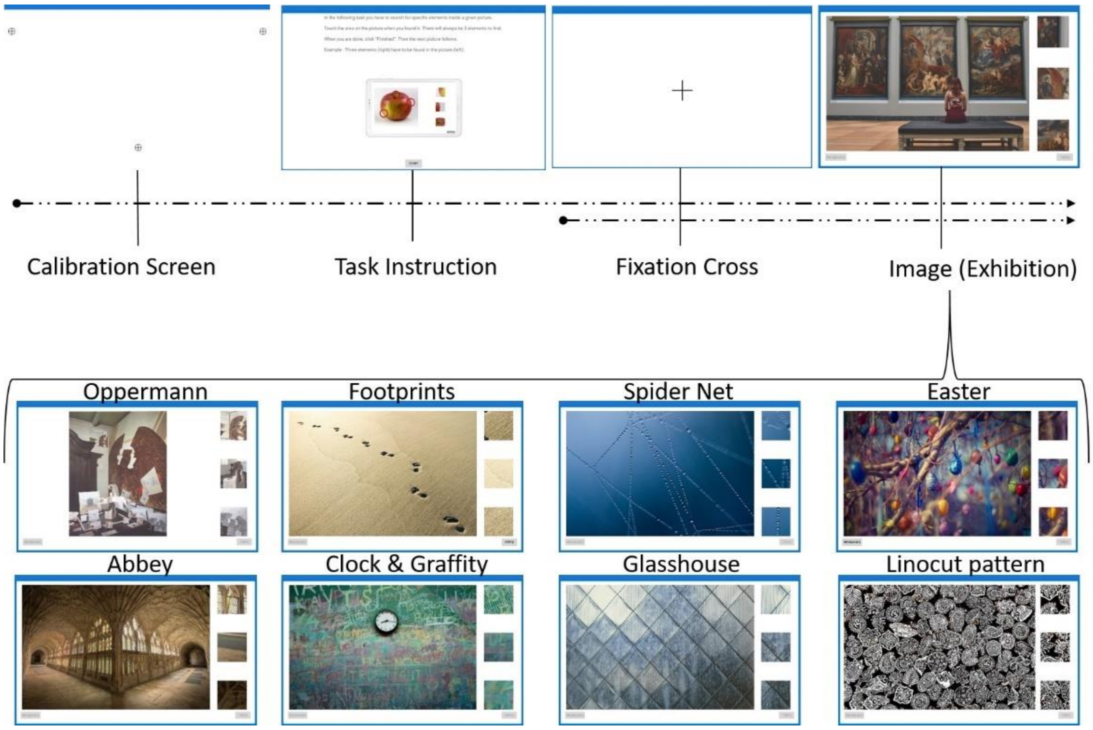
Sample II comprised 52 participants, who were screened for eligibility as part of the eye-tracking study. Two participants were excluded from further analysis, one because of poor acuity and the other because of insufficient eye-tracking quality. For another participant in the expert group the eye-tracker lost the tracking signal on two trials and therefore data from this participant were only included for the remaining trials (images 1-7). As there is currently no validated test available on the assessment of VL, experts and novices were screened based on their prior experience and interest in the visual arts. Participants in the expert group (n= 25) were either members of the European Network of Visual Literacy (ENViL) or working in professions requiring a high visual competency (photographer, gallerist, art educator, art designer, art students, or self-employed artists). The non-expert group (n= 25) were adults from the clerical and academic staff of various educational settings not associated with academic or professional work in the visual arts. The participants' ages ranged from 16 to 66 years (*mean* age = 29.08 years, *SD* =12.55 years). Participants in Sample II were individually assessed in seminar or laboratory rooms (e.g. at the Academy of Fine Arts in Munich) or at expert's working places. All participants had normal or corrected-to-normal vision. Student participants received 20€ each as compensation. Other participants, including experts in the expert group, who were generally interested in the topic of visual literacy and eye tracking, participated without any compensation.

Stimuli

Subjects were required to identify three targets (details) on each of the four (Experiment I) or nine (Experiment II) subsequently presented images. Figure 1 illustrates the procedure of the experiment. Calibration screen and fixation cross was only visible for Experiment II.

Figure 1

Procedure of experiment with nine images



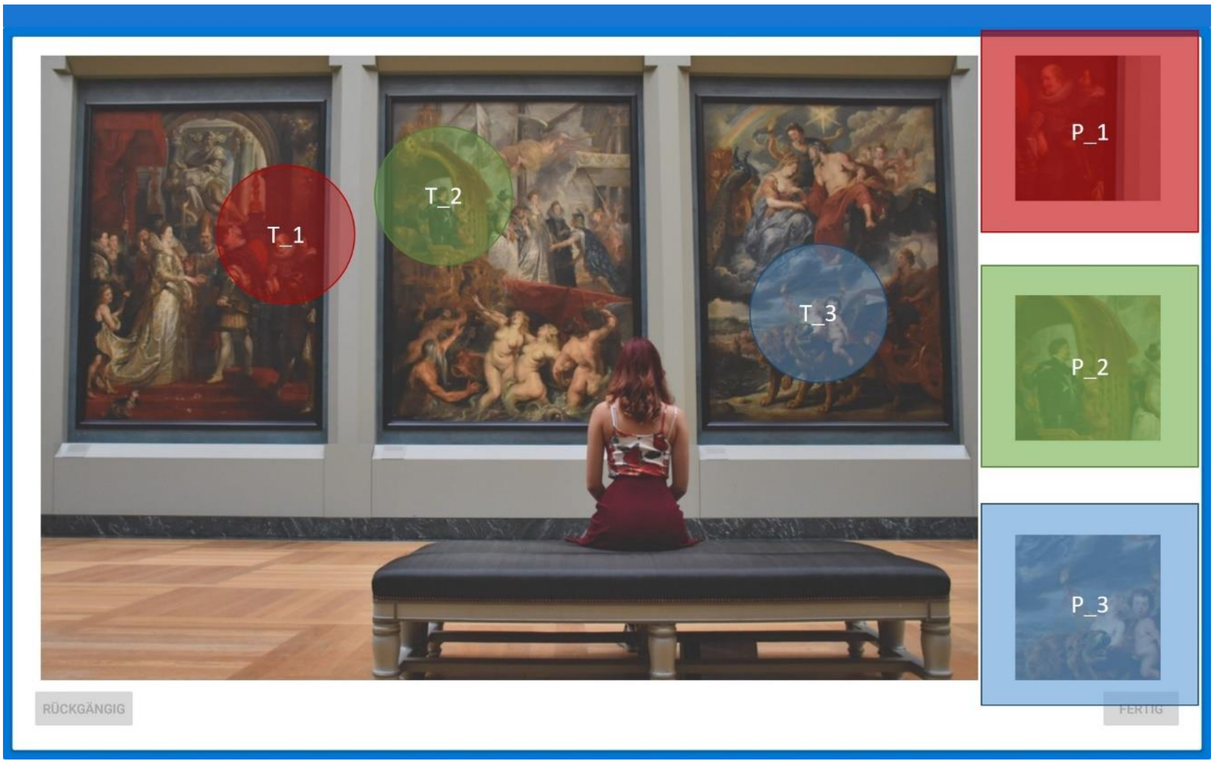
Note. Sample I only included the following images: Exhibition, Oppermann, Footprints and Clock & Graffiti. Sample II included all nine images.

In a short pre-study, four untrained VL experts independently rated each of the photographs in our sample of images with the ARESI classification (van Meel-Jansen, 2006) on a scale from 1 (feature not present) to 7 (highly prominent feature). Ratings reached a reasonable mean interrater correlation of ICC = 0.50 (Shrout & Fleiss, 1979) and the following mean values: A= 4.8, R= 5.1, E= 3.6, S= 5.1, I= 4.4. Thus, images can be regarded as of satisfactory aesthetic value with above average rating on realism (R) and style and form (S) and not representing outliers on one of the five aesthetic domains.

Each image had three primes positioned on the right-hand side (P_1, P_2, and P_3). The primes represented details to be found as targets on the left (T_1 to T_3). Figure 2 shows prime and associated targets as pre-defined AOIs (see Appendix B for all nine images). Image areas not covered by any AOI are defined as white space (WS). Once the subject found (or thought to have found) a target they touched the region on the screen to indicate the corresponding position of the given target. A red circle of 50-pixel radius appeared at each touching point to indicate that input was registered. Targets were counted as identified when they were touched within a 50-pixel radius around the center of each target region.

Figure 2

Example image "Exhibition" with pre-defined AOIs



Note. Primes (P_1 to P_3) on the right-hand side and targets (T_1 to T3) of the image are correspondingly shaded in color. Note that neither the colored shading nor the labelling of AOIs was presented to the subjects during the experiments. See supplementary material for all nine images (Appendix B).

All participants were assessed on Android A6 Tablets with 10.1-inch screen size. Tasks were constructed explicitly for the study (Andrews et al., 2018). There was no time constraint during the task. The participant ended each trial by pressing the "Done"-button.

In Sample II eye movements were recorded with SMI eye-tracking glasses (SMI ETG 2w Analysis Pro). The glasses were positioned and strapped tightly onto the subject's head, which they could freely move during task completion. Participants were seated 50–80 cm away from the tablet screen. Eye movements where calibrated with a 3-point calibration. All eye-tracking data were recorded at 60 Hz. Saccades and fixations (as well as blinks) were recorded binocularly. Before each image was presented, a fixation cross was displayed for 2 seconds. Subjects were free to search the targets in any order and received no further feedback during the trial (on number of correct targets found).

The session started with a task instruction (in German):

"In the following task you have to search for specific details inside a given picture. Touch the area on the picture where you found it. There will always be 3 details to find. When you are done, click 'Done'. Then the next picture follows."

Eye-tracking data analysis was conducted with SMI BeGaze^{TM} version 3.7. Fixations for each image were mapped onto corresponding reference images using SMI fixation-by-fixation semantic gaze mapping (Vansteenkiste et al., 2015). Each reference image was divided into three prime and three target AOIs (see Fig. 2). The following eye movement variables were analyzed: the spatial coordinates of each fixation, the fixation sequence and the fixation duration in milliseconds. Due to the explorative nature of this study, no measures against inflation of type I error were undertaken, as statistical tests were not regarded as confirmatory analyses.

The study was conducted according to the guidelines for human research outlined by the Declaration of Helsinki and was approved by the Ethics Committee of Research of the Leibniz Institute for Research and Information in Education, Frankfurt am Main (DIPF, 01JK1606A). All subjects and their legal representatives respectively had given written informed consent prior to participation.

Latent Profile Analysis

Students' responses to the images presented were recorded as a vector of 4 (images) times 3 (details to be identified) = 12 dichotomous variables (target correctly identified or not?) and 4 continuous variables (time in sec. to solve all three search tasks per image). Individual response patterns were grouped into latent classes of similar response patterns by means of a Latent Profile Analysis (see Ferguson et al., 2020), for statistical model and a practical application). Models between 2 and 6 latent classes were estimated using MPLUS 8.4 software. The decision to interpret four latent classes (named LC1-LC4 in Figure 4) as the final solution was based on the progression in the BIC fit indices (sharp decline after 4 classes) and the Lo-Mendell-Rubin test of significant improvements in model fit ($p = 0.6707$ for a five class solution). Latent class analysis (and the generalization of latent profile analysis) results in class membership probabilities for each individual to each of the estimated classes. Individual students were manifestly classified into latent classes according to their modal class membership probability. This categorical variable (reflecting four qualitatively differing solution patterns) then was used as dependent variable, which was regressed on by a list of demographic (gender, age) variables and self-reported skills (mentioned above under "subjects"). To arrive at a parsimonious multinomial logistic regression model for group membership, a stepwise selection of predictor variables was applied, which resulted in three significant predictors (Table 2). Results are presented as Odds-Ratios per scale point of three self-ratings of students' visual performance skills for three of the latent classes as compared to the largest ("mainstream") group.

Hidden Markov Model

Hidden Markov models (HMMs) represent efficient and flexible modelling tools for data that include temporal constraints and spatial variability such as the sequences of eye movements. The intuitive idea behind a Markov Model or a Markov chain is that in series of events where each probability of something happening depends only on what happened right before it. For eye movements, we can look at the fixation sequence and classify each fixation to their most likely state (data-driven Area of Interest) depending on the previous fixation. The HMM divides the image into multiple data-driven AOIs which we can call Markov *states*. Each time a new fixation arises in our fixation sequence, we can give the fixation a certain probability of belonging to the same AOI or switching to any other one. This probability is called the *transition probability*. The transition probability is conditional to the previous fixation observed. Combining the probabilities for each state gives us a *transition matrix*. In the case of *hidden* Markov models the states are not directly observed. Only the observation sequence (the fixation sequence) is known.

A HMM comprises three components: the initial state distribution (in what states participants start in), the state transition probability distribution (how likely it is to transition from one state to another), and the observation probability distribution (how likely an observation is produced by any given state, i.e. how likely a fixation is linked to any given AOI). Again, each HMM state in this study represents a location on the image participants fixated while inspecting that area. The transitions between each hidden state (image area) can be placed into a transition probability matrix, describing the probability of switching between each hidden state.

A Hidden Markov model can be defined as

$$\boldsymbol{\lambda} = \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi}$$

where λ is a triplet comprising the model matrices. A is the state transition probability distribution of state j following state i. B is the observation (emission) probability distribution of observation k from the state j. π is the initial state distribution: $\pi = \{\pi i\},\ 1 \leq i \leq N$.

See (Rabiner, 1989) for an introduction to HMM and (Coutrot et al., 2018) and (Boccignone, 2019) for HMM applied to eye movement data.

HMMs were estimated using the depmixS4 package (Visser & Speekenbrink, 2010) with the software R. Spatial variability can be modelled through the output distribution of an HMM and temporal variability through the HMM transition parameters. In the case of eye-tracking data, each transition can be interpreted as an outgoing saccade from one data-driven AOI to another. The HMMs were formulated based on the spatial coordinates of each fixation. No additional constraints were put on the models' parameter matrices. We estimated HMMs for each expertise group (experts vs. non-experts) on each of the nine images.

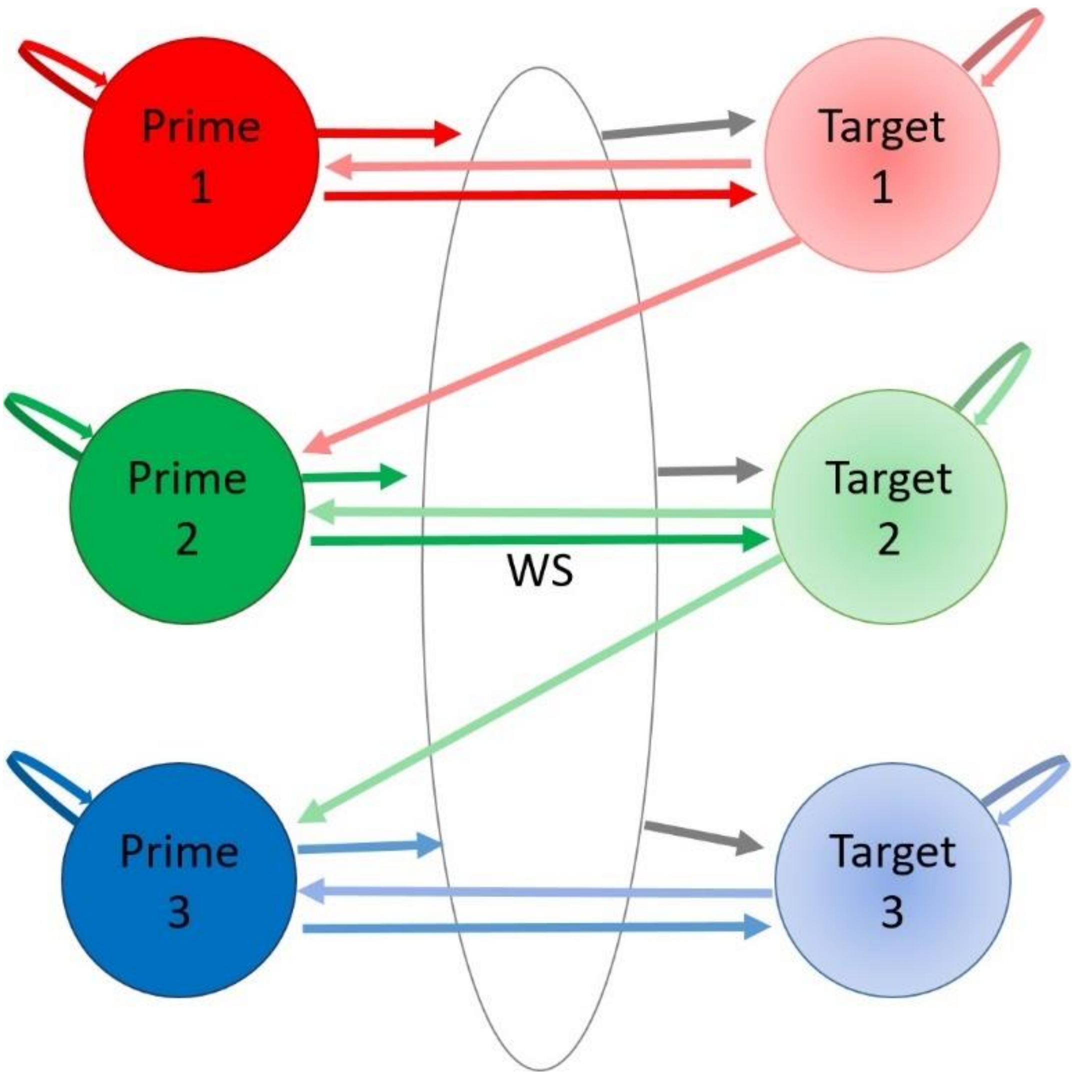
Each HMM was estimated from two to 14 states. Selection of model class (number of states for each image and group) was achieved by Bayesian Information Criteria (BIC, see Vrieze, 2012). If there was a discontinuous progression of the log-likelihood when incorporating a new state to the model, alternative seeds to determine randomly chosen starting values were used to avoid local optima.

To visualize the HMM, all fixations (emission points) were exhaustively and disjunctively classified to their best suited hidden Markov state and then drawn as 2D density maps (contour maps) onto the corresponding images. In order to analyze the precision of the visual search (H3) each pre-defined AOI (primes and targets) was linked to the hidden state with the highest number of fixations. Percentage of fixations inside pre-defined AOIs in each corresponding hidden state was used to determine precision. High fixation overlap of hidden state and AOI indicates higher precision while looking for prime and target regions.

Figure 3 visualizes a hypothetical 7-state HMM of plausible transition probabilities between primes and targets.

Figure 3

Hypothetical Transition Probability Matrix for a Theoretical HMM



Note. In this simple arrangement each AOI represents a hidden state (Prime 1 to 3, Target 1 to 3 or White Space). Each hidden state has a certain probability of staying in that state or transitioning from one to another indicated by arrows. Note that the pairwise numbering is arbitrary as there was no instruction to search from top-to-bottom.

3 Results

Speed and precision of search

Table 1 presents error rates for each target and task durations on each of the four images presented in Sample I and compares it to the corresponding results for Sample II. In Experiment I (Sample I) only images Exhibition, Oppermann, Footprints, and Clock & Graffiti were shown.

Table 1

Percent of Correctly Solved Targets and Mean Time to Solve each Image in Sample I and II

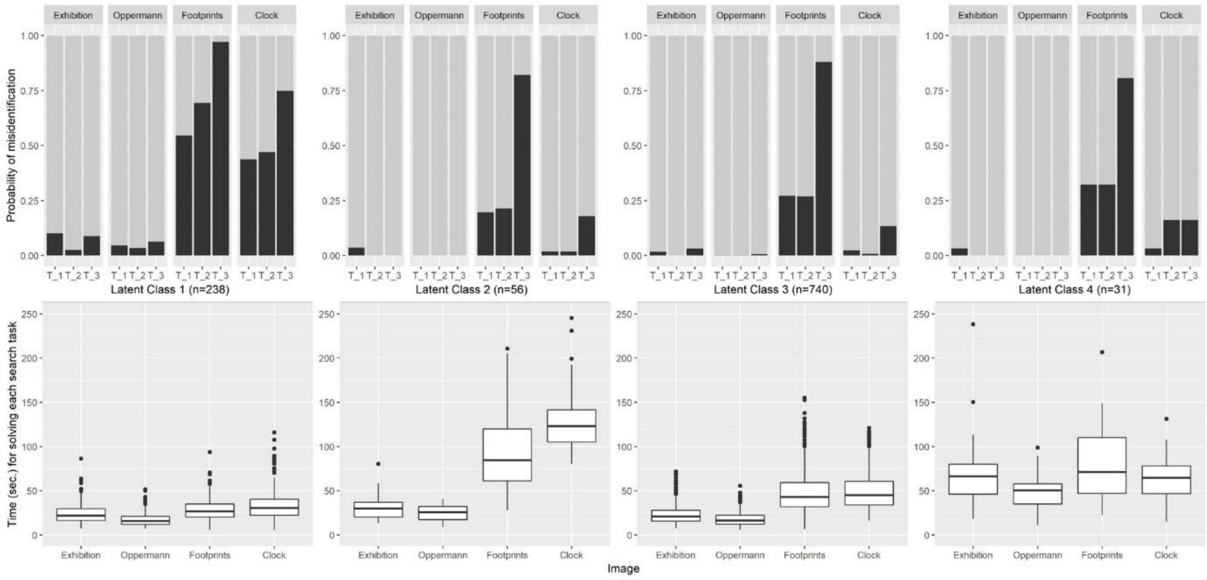
Task		Total Sample I (N=1056) Mean age= 15.27 years (SD= 0.94)		Sample II Experts (N=25) Mean age= 34.36 years (SD=14.69)		Sample II Non-experts (N=25) Mean age= 23.80 years (SD= 6.92)	
Image	Target No.	Error Rate	Mean time in seconds (SD)	Error Rate	Mean time in seconds (SD)	Error Rate	Mean time in seconds (SD)
Exhibition	1	0.038	25.29 (15.14)	0.20	20.52 (6.58)	0.08	18.81 (6.02)
	2	0.006		0.00		0.00	
	3	0.041		0.04		0.04	
Opper-mann	1	0.011	19.16 (10.10)	0.04	15.07 (6.77)	0.00	13.87 (7.15)
	2	0.008		0.00		0.04	
	3	0.018		0.00		0.00	
Footprints	1	0.331	47.21 (29.54)	0.32	44.06 (17.86)	0.20	48.73 (33.81)
	2	0.362		0.28		0.48	
	3	0.895		0.76		0.88	
Clock & Graffiti	1	0.115	50.39 (28.96)	0.12	55.18 (36.43)	0.12	47.63 (23.94)
	2	0.115		0.00		0.08	
	3	0.274		0.16		0.20	
Spider Net	1	Not applicable		0.56	41.25 (18.72)	0.68	35.41 (12.03)
	2			0.08		0.08	
	3			0.20		0.32	
Easter	1	Not applicable		0.12	37.29 (16.17)	0.16	33.73 (17.15)
	2			0.08		0.24	
	3			0.00		0.04	
Abbey	1	Not applicable		0.04	47.86 (25.00)	0.00	51.55 (20.00)
	2			0.04		0.00	
	3			0.20		0.60	
Glass-house	1	Not applicable		0.12	60.12 (38.7)	0.28	57.90 (30.23)
	2			0.40		0.40	
	3			0.20		0.20	
Linocut Pattern	1	Not applicable		0.08	86.37 (40.4)	0.00	105.52 (62.69)
	2			0.16		0.24	
	3			0.04		0.08	

Students were able to correctly identify the required targets virtually without errors for the Exhibition and the Oppermann image (Table 1) and solved the search task in mean durations of about 20 to 25 seconds. The Clock & Graffiti image comprised a lot of optical distractors and therefore led to an error rate of at least 11.5% per target. Targets in the Footprint image were much harder to identify with at least one third of all students failing per target. The two more difficult images required on average double the task duration (about 50 seconds) as compared to the other two images. Remarkably, the third target of each of the four images was the most difficult one for all four tasks. Footprint target 3 was only correctly solved by slightly more than 10% of the students. VL experts, but also novices on average solved the easy images faster than students, with the exception of the Clock & Graffiti image, where VL experts worked longer than students (H1a). Experts were as good or better at identifying targets in comparison to the student sample (H1b). Even though experts found on average one target more than non-experts M_{Expert} = 22.76 (SD=1.69), $M_{Non\text{-}experts}$= 21.56 (SD=2.65), this was not statistically significant; two-sided Welch t(40.781)=1.91, p=.063 (H2a). Across all 9 images experts did not differ from non-experts with respect to time on task (M_{Expert} = 45.30 sec., $M_{Non\text{-}experts}$= 45.91 sec, F(1,48)= 0.022, n.s. (H2b).

When error patterns over all four images and invested time periods were grouped into latent classes of similar behavior, a latent profile analysis resulted in four distinguishable patterns (see methods section for details justifying the decision for 4 classes). Latent class 3 (LC3, n=740) more or less represents the same solution pattern (error rates, durations) as the total sample with the exception of Clock & Graffiti, where LC3 performed better than the average. Errors cumulate in the second largest class LC1 (n=238), where students performed reasonably on the Exhibition and Oppermann image, but failed to identify targets over base rates of the Clock & Graffiti and the Footprint image. The reason for this low performance might be given by the high task-performance speed that members of LC1 displayed especially for the more challenging images. The remaining quite small groups (LC2 and LC4) differ mostly with respect to the time invested for solving the search tasks. LC2 (n= 56) worked fast on the two easy images (and achieved nearly perfect hit rates), but invested much more time (96 sec. and 125 sec.) for the more difficult images. By doing so, they were able to achieve hit rates comparable to or better than the “mainstream group” LC3. By contrast, members of LC4 (n=31) represent a group that continuously worked quite slowly (all mean times above 50 sec.) over all four images, but ending up in error rates not better than the mainstream.

Figure 4

Probability of Target Misidentification for each Latent Class on each Image (upper Part) and Time for Solving each Search Task on each Image (lower Part).



Note. Latent class profiles are depicted from left to right: LC1 “fast and erroneous“ (n=238), LC2 "easy images fast, difficult images slowly" (n=56), LC3 “Mainstream” (n=740), and LC4 “slow working" (n=31).

A multinomial logistic regression model on the solution pattern as represented by class membership explored the potential impact of gender, age, and metacognitive self-perceptions of students in Sample I. Only three variables reached a nominal significance level of $p < 0.05$. Gender and age did not affect solution patterns, neither did the variables “Art is important for me personally”, “When I see a photographed geometric object, I can imagine what it looks like from behind”, “I can remember small details in pictures”, “I like to solve picture puzzles”, “Do you regularly attend an art school or art workshops?”, “Understanding art history and theory”. But the three variables listed in Table 2 had a significant impact on class membership. The global Likelihood Ratio Test for the whole model scored at $\chi^2_{(9)} = 64.903$ ($p < 0.001$) and resulted in a Pseudo-R-square of 0.0746. Each of the three regressor variables reached a Wald Chi-square test with $p < 0.001$. Specific effect sizes (Odds Ratio per increasing response category) of the independent variables on the probability of each solution pattern (LC4, LC2 and LC1 compared to the mainstream pattern LC3) are listed in Table 2.

Students claiming to have a photographic memory display a lower probability for belonging to each of the three non-mainstream latent classes at each increased response category. Most pronounced is this effect for LC1 (“fast and erroneous”). Students belonging to this group (on average) proclaim to have a photographic memory to a smaller degree. If students are convinced of their ability to visualize things mentally, then this slightly diminishes their chances to belong to latent classes LC4 and LC2, but increases the probability for membership in LC1 (“fast and erroneous”) by more than 50% per category. This means, that LC1 has a self-perception of high competence in recognizing details in pictures and might therefore work very fast on the respective tasks, but indeed fail to reach the same precision as the other groups. Students’ self-reported high performance in art education is associated with higher chances to belong to the “slow working” group LC4, but considerably lower chances to belong to LC2 or LC1.

Table 2

Effect sizes (Odds Ratios) of self-reported art skills on Latent Profile classification

Effect	Comparison Group (Reference = LC3 "mainstream")	Odds Ratio (per category)	95% Confidence Limits	
“I have a 'photographic memory”	LC4 "slow working"	0.961	0.632	1.463
	LC2 "easy images fast, difficult images slowly"	0.928	0.671	1.284
	LC1 "fast and erroneous"	0.659	0.554	0.785
“I can visualize things mentally”	LC4 "slow working"	0.991	0.659	1.491
	LC2 "easy images fast, difficult images slowly"	0.992	0.716	1.375
	LC1 "fast and erroneous"	1.536	1.298	1.817
“How well do you perform in arts education generally?”	LC4 "slow working"	1.057	0.686	1.629
	LC2 "easy images fast, difficult images slowly"	0.591	0.435	0.802
	LC1 "fast and erroneous"	0.721	0.605	0.860

From these results it seems clear that interpreting a simple score of correctly solved search tasks does not cover art related visual competence in a meaningful way. A deeper understanding of the cognitive processes during search tasks has to be acquired from additional images and from comparing VL experts to VL non-experts. Results from sample II might contribute to this understanding.

Eye-movements in Sample II

Table 3 shows the mean fixation duration and mean number of fixations on each image. VL experts generally show longer fixation durations than non-experts. Experts' mean fixation durations ranged from 264.83ms ("Linocut Pattern") to 317.59 ("Abbey") and non-experts' mean fixation duration ranged from 256.26 ms ("Linocut Pattern") to 308.12ms ("Abbey"). For most images, experts exhibited more fixations than non-experts. As there is a significant difference in age between both expertise groups ($t(34.163)=3.252$, $p<.01$ with $M_{Experts} = 34.36$, $M_{Non\text{-}experts}= 23.80$), a correlation between age and eye movement indicators was calculated for possible confounders. Age and fixation duration exhibited a moderate correlation of $r=0.37$, $p<.01$ and no correlation between age and overall number of fixations was found ($r=0.08$, n.s).

Table 3

Mean Fixation Duration (ms) and Number of Fixations per Image

Mean fixation duration (ms)

Image	Experts (N=25)	SD	Non-Experts (N=25)	SD
Exhibition	302.35	260.12	287.94	249.74
Oppermann	292.27	230.77	289.26	227.65
Footprints	284.29	195.00	265.24	182.67
Spider Net	280.09	228.07	288.71	221.24
Easter	274.69	201.86	258.32	180.26
Abbey	317.59	227.79	308.12	222.22
Clock & Graffiti	297.96	191.01	287.28	197.67
Glasshouse*	313.90	219.25	292.40	201.58
Linocut Pattern*	264.83	148.37	256.26	144.71

Mean number of fixations

Image	Experts (N=25)	SD	Non-Experts (N=25)	SD
Exhibition	59.83	15.76	54.56	19.24
Oppermann	43.83	18.92	40.24	21.19
Footprints	129.46	48.21	154.00	117.97
Spider Net	136.67	55.17	104.84	33.48
Easter	117.17	43.13	108.92	60.41
Abbey	136.42	64.07	142.40	47.16
Clock & Graffiti	168.62	92.71	141.04	71.24
Glasshouse*	177.22	107.66	162.48	85.99
Linocut Pattern*	298.78	110.99	346.80	205.27

Note. *N=24 experts due to insufficient data transmission from eye-tracker during one session.

HMM Estimates

Table 4 shows the optimal number of hidden states based on HMM by expertise group, i.e. fixation sequences of all experts are used to estimate one HMM for the entire expert group and all fixation sequences of all non-experts for the non-expert group. There is only a small difference in the number of states in the expert group and in the non-expert group ($Mean_{Experts}$=8.7 hidden states vs. $Mean_{Non\text{-}Experts}$=7.8 states) that varies depending on given image.

Table 4

Optimal Number of Hidden States based on HMM by Status Group*

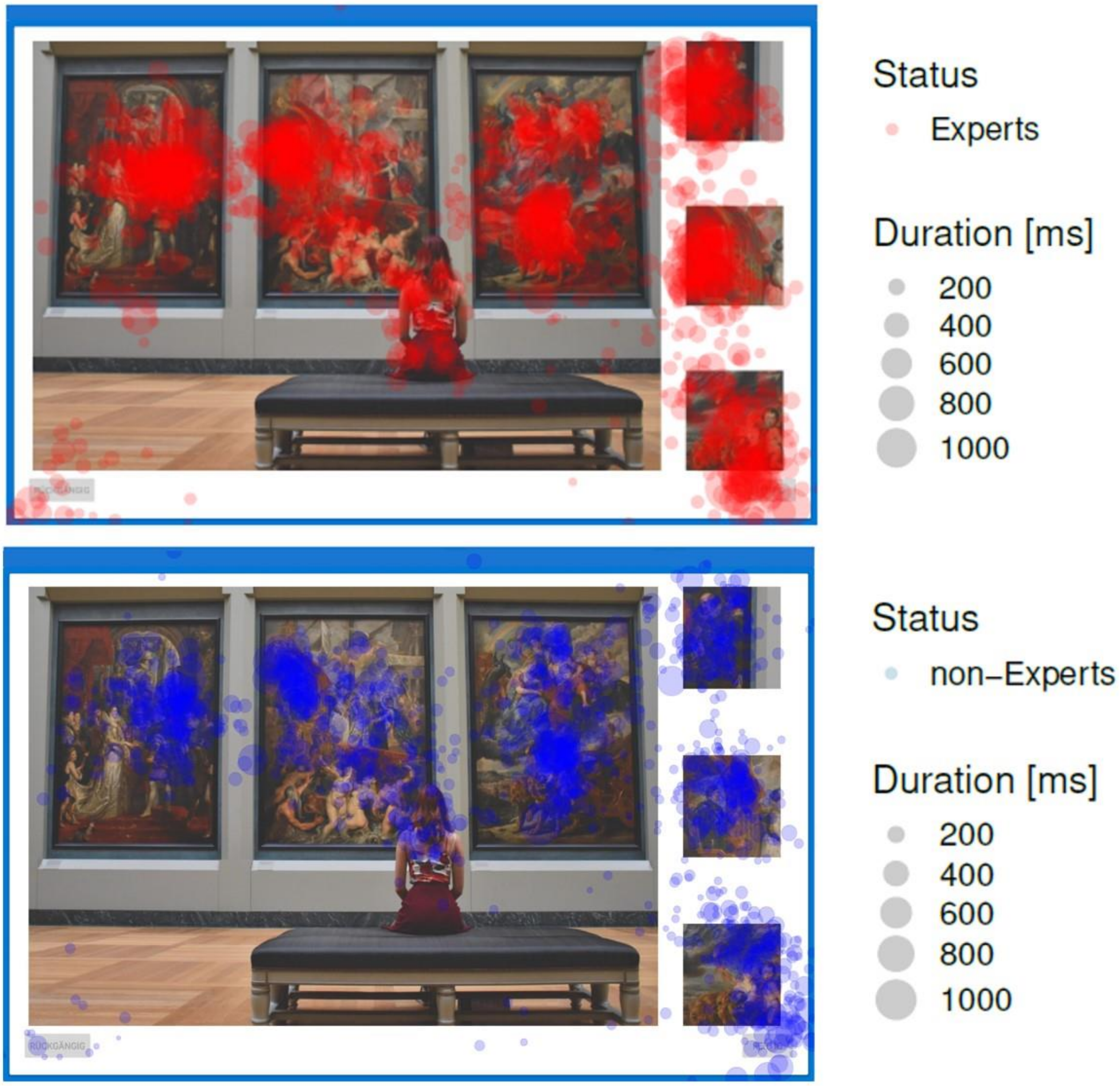
Image	Expert Group (N=25)	Non-Expert group (N=25)
Exhibition	9	9
Oppermann	7	8
Footprints	9	6
Spider Net	10	7
Easter	7	7
Abbey	8	10
Clock & Graffiti	8	8
Glasshouse**	10	9
Linocut Pattern**	11	7

Note. *as indicated by Bayesian Information Criterion (BIC), **N=24 Experts due to insufficient data transmission from eye-tracker during one session.

A traditional Heatmap (Bojko, 2009) of fixations and fixation durations of Experts and non-Experts is given on the image Exhibition (Figure 5). As can be seen, fixations tend to cluster around prime and target image regions. Also some fixated areas seem to overlap each other. What about the fixations in between, i.e. to WS? If we want to classify each fixation to their respective underlying image region, we can use the HMM to infer the most probable state for each fixation. We also take the sequence of fixations, as a transition probability matrix into account while modelling (in contrast to other classification methods like k-means clustering (Steinley, 2006)).

Figure 5

Heatmap of fixation points of experts (red) and non-experts (blue) on image Exhibition

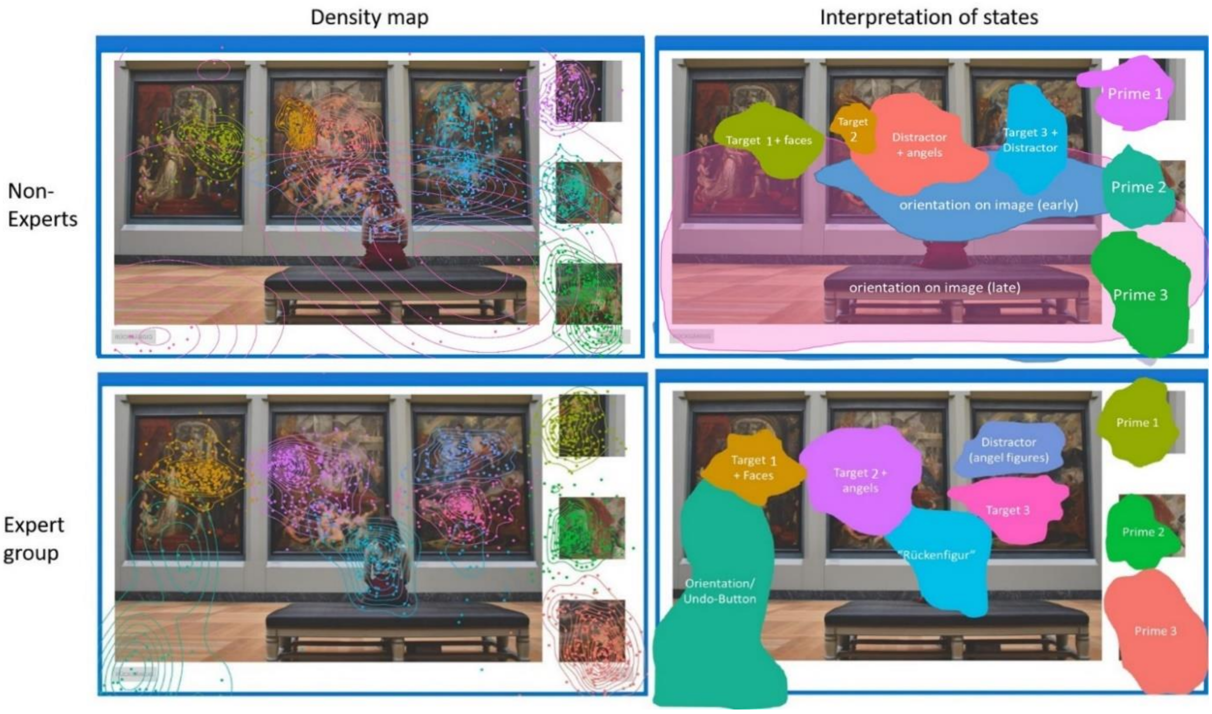


Note. Each circle represents a fixation. Size of circles represents fixation duration.

As each HMM is based on spatial coordinates of fixation points and their sequence, hidden states can be analogously visualized as 2D topological areas on the image incorporating additional information. Figure 6 represents hidden areas for the non-expert group and expert group on image Exhibition (see Appendix B for all nine images). Every hidden area was assigned a different label, either belonging to a prime, target or distractor area (part of WS). Interestingly, the novice group includes a broader orientation area which spans across the image. We can differentiate this clearly now, as each fixation is matched to their most probable state over time, allowing for spatially overlapping states differentiated additionally through the transition probability matrix. By contrast, the expert group clearly defines a figure of a woman in front of the museum paintings, resembling the traditional motif of a "Rückenfigur".

Figure 6

HMM Density Map of Hidden States based on the Fixation Sequence for each Group (Left) and Assigned Semantic Interpretation of the Hidden States (Right)

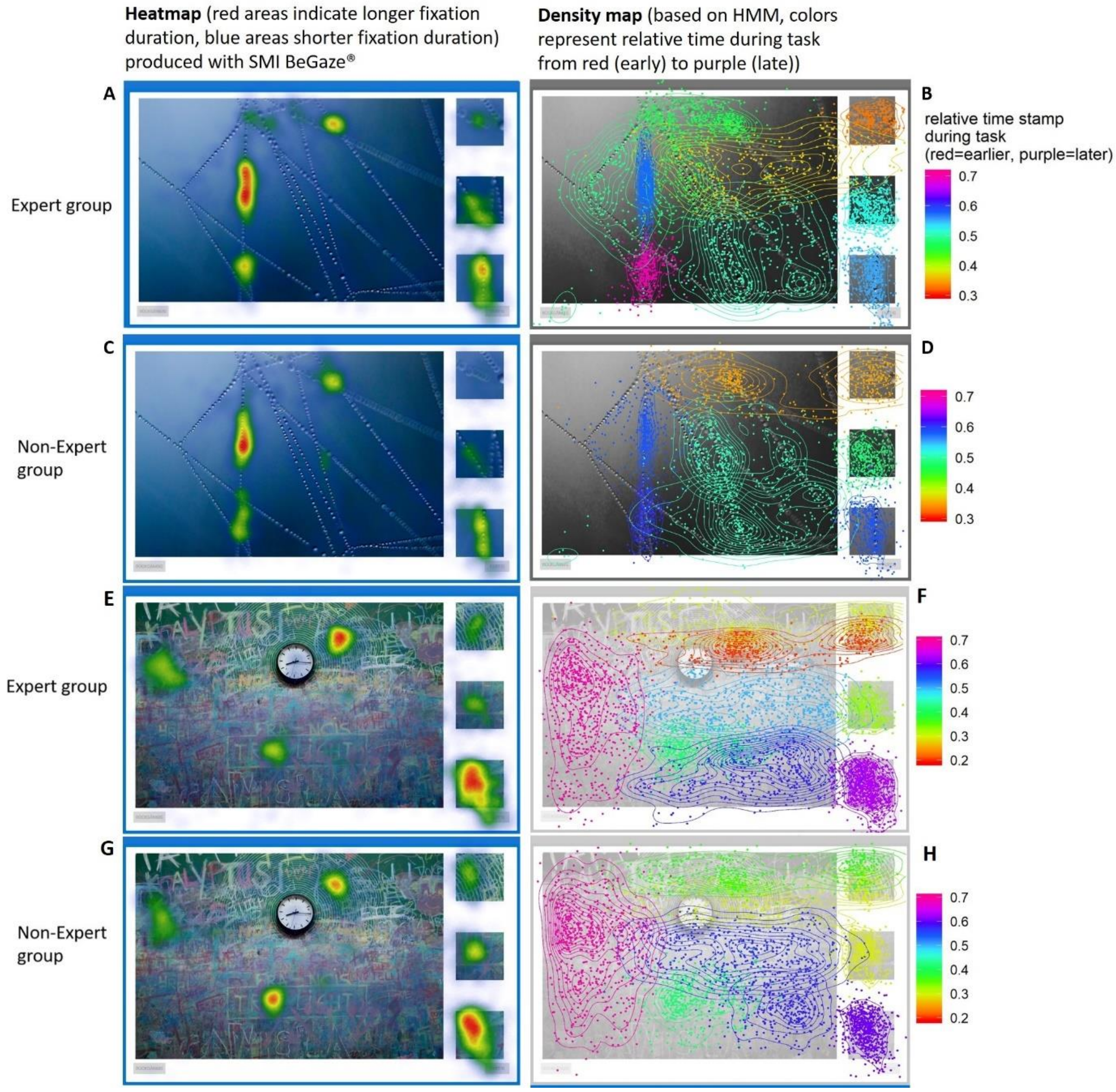


Note. States are colored for better differentiation. Each hidden state either represents a prime, target or distractor region in WS. The nine states of the non-expert group are: (1) Prime 1, (2) Prime 2, (3) Prime 3, (4) Target 1, (5) Target 2 + faces, (6) Target 3 + Distractor, (7) Distractor + angels, and (8, 9) as two wide orientation states across the image. The nine states of the expert group are: (1) Prime 1, (2) Prime 2, (3) Prime 3, (4) Target 1 + angles, (5) Target 2 + faces, (6) Target 3, (7) angel figures as distractor, (8) "Rückenfigur", (9) orientation/undo-buttons. See supplementary material for all nine images (Appendix B).

Figure 7 compares another conventional Heatmap (produced by SMI BeGaze) and a density map based on the HMM. Regions attracting attention during different phases of the visual search become visible through HMM states, otherwise they are overlooked in conventional Heatmaps. For example prime 1 (top right corner) is hardly evident by the Heatmap. However, the HMM density map shows us how prime 1 is connected to the target region (at the top) and incorporated into a common state. In this case, experts as well as non-experts seem to fixate two pre-defined AOIs in rapid succession to form a single HMM state. Both examples given in Figure 6 and 7 show how regions not previously identified (in WS) that attract participants' attention are uncovered with the use of HMMs. Figure 7 additionally shows how overlapping hidden states differ in their relative time (normalized to each subject's individual working time).

Figure 7

Heatmaps vs. Density Maps on images ("Spider Net" and "Clock and Graffiti") for the expert group and non-expert group

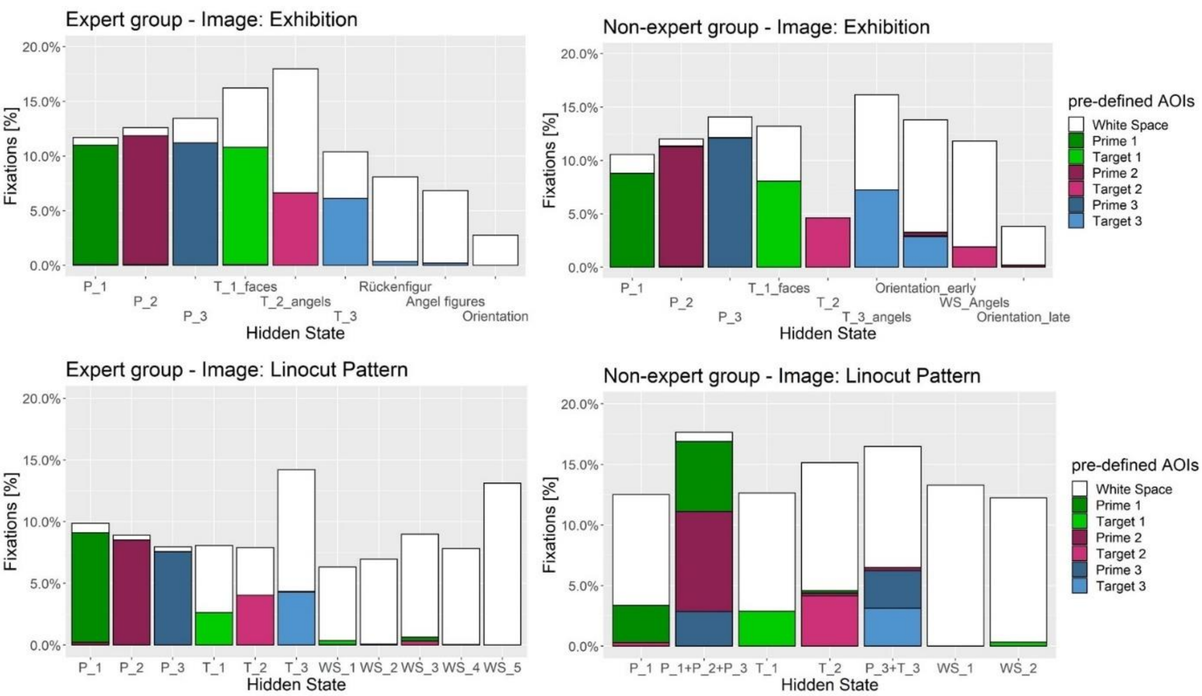


Note. The Heatmaps (left; A, C, E, G) disregard fixation points that are scattered outside of "hot spots" because on a global scale they do not carry much information. Image underlying density maps are masked out to make colors more visible. The HMM density maps (right; B, D, F, H) shows how every fixation point is connected to their most probable state throughout the image (color indicates differences in relative time of attendance). Some hidden states span over multiple areas or overlap each other that are still differentiable through time. We can also see how local differences in density differ within each state.

How good does each hidden state represent pre-defined AOIs? Figure 8 displays the overlap of hidden states with pre-defined AOIs. There is a clear connection between the hidden states and the pre-defined AOIs. In most cases one hidden state is directly associated with one prime or target AOI, indicating a meaningful distribution of hidden areas on the image. The more “white” a hidden state encompasses the more it includes undefined WS region on the image. Usually the target AOIs include more WS. Some hidden states in WS can be defined as distractor areas that draw attention during the search. When a hidden area covers multiple pre-defined AOIs (see Fig. 8 on image “Linocut Pattern” below, e.g. in state 2 for non-experts) they combine into a single state. This may happen due to frequent transitioning between two pre-defined AOIs. We can see how the states are not randomly distributed over the image, but are closely related to the visual search task (primes and targets).

Figure 8

Distribution of Fixations and Percentage Overlap on Pre-defined AOI and Hidden States for the Expert (left) and Non-Expert (right) group



Note. Each color represents either a target or prime region on the image. White is undefined White Space (WS) area. In most cases prime regions are represented by a single hidden state; e.g. states for target and prime regions in the expert group on image Exhibition. Target regions are usually encompassed by WS area, as they need to be found inside the image. Frequent transitioning between multiple pre-defined AOI may integrate into one hidden state; for example in the non-expert group on image "Linocut Pattern" (bottom right) one hidden state overlaps with P_1, P_2 and P_3.

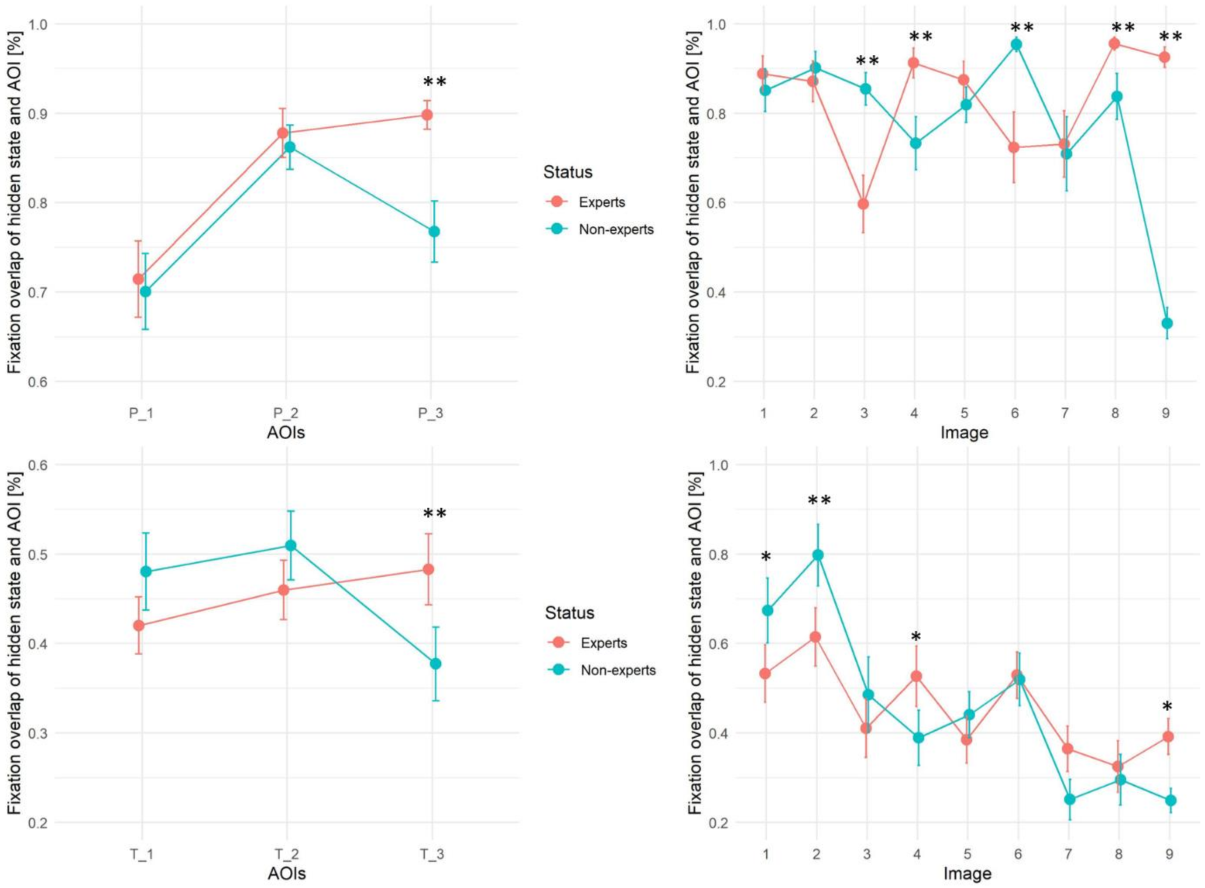
In order to determine differences in precision between experts and non-experts (H3), hidden states in each HMM were assigned to 3 primes and 3 target AOIs according to their best possible fit; each pre-defined AOI to the HMM-state with the most fixation overlap, i.e. the percentage of fixations of the corresponding HMM that fell into the predefined AOI region, (see Fig. 8). A two-way, repeated measures (three primes and three targets as within factor 1, nine images as within factor 2) ANOVA with one between-subjects factor (VL-experts vs. non-experts) was estimated using these percentages as dependent variable measuring the precision of the fixations, one ANOVA model for primes, one for targets. No main effect for experts vs. non-experts could be seen in either of the models: though the global F-test for primes pointed at differences ($F_{primes}(1,48) = 13.25$), the global F-test for targets ($F_{targets}(1,48) = 3.96$) fell short to reach significance and no clear direction of differences was visible (see Figure 9). For both primes and targets there was a significant main effect for images ($F_{primes}(8,41) = 182.77$; $p < 0.001$; $F_{targets}(8,41) = 67.33$; $p < 0.001$), and for the three AOIs ($F_{primes}(2,47) = 105.38$; $p < 0.001$; $F_{targets}(2,47) = 8.84$; $p = 0.0006$).

Both within-subjects effects interacted with the status of experts vs. non-experts in the following way: the third prime and target AOI (P_3 and T_3) on all 9 images was on average fixated by experts with a significantly higher precision than by non-experts (see Figure 9). Non-experts on P_3 and T_3 more frequently switched between prime and target and WS regions to reassure that they recognized the correct region on the image, whereas VL-experts once they had memorized the third prime (coverage rate nearly 90% vs. 75%), VL-experts were able to search the corresponding target quite efficiently (coverage rate nearly 50% as compared to only 37% in non-experts). With regard to differences between images, VL-experts reached a significant higher precision in fixating the primes on the "Spider Net", "Glasshouse" and "Linocut Pattern" image, and in fixating the targets of the "Spider Net" and "Linocut Pattern" image. Non-experts on the other hand more precisely fixated the primes of the "Footprints" and

"Abbey" image indicating that they spent more fixations within the prime regions and reached higher precision in fixating target regions in image "Exhibition" and "Oppermann".

Figure 9

Fixation Overlap between pre-defined AOIs and Best-Fitting Hidden State (State with Highest Fixation Overlap) for Expert and Non-expert Group.



Note. Upper panels (A, B) show overlap between prime regions, lower panels (C, D) between target regions. Images: 1=Exhibition, 2=Oppermann, 3=Footprints, 4=Spider Net, 5=Easter, 6=Abbey, 7=Clock & Graffiti, 8=Glasshouse, 9=Linocut Pattern. ***p*<.001, **p*<.05.

4 Discussion

Student solution patterns

The majority of students (n=740) took their time to solve the more difficult items "Footprints" and "Clock and Graffiti". Most of these "mainstream" students claimed to have a 'photographic memory' to a higher degree than other student groups. Overconfidence in student's ability to "visualize images mentally" increases the probability of them rushing through the tasks and making more mistakes. Visualizing images mentally can go beyond details in 2-dimensional images and therefore not be helpful for the search task. A "photographic memory" much more encompasses the necessary skill to attentively focus on details in artwork. On the contrary, high self-reported capacity in art education leads students to take it slower on all tasks but does not necessarily enable students to outperform other student groups. Art teachers therefore may be interested not only in students engagement in art class but also be inclined to know about visual memory ("photographic memory") skills. These results can be first clues for finding student groups that need more help in engaging and analyzing artwork or to learn when to invest more time in a visual task.

VL experts could solve three of the common four images faster than students and in a more homogeneous manner (less variance of solution time) (H1b) and at equal or even superior chances for a correct solution (H1a). Only for the "Clock & Graffiti" image did VL experts take longer than students, resulting in a nearly perfect solution probability that was not reached by students.

Expertise differences

Did VL experts outperform students? This cannot be deduced from mere test solutions, but requires data on the solution process as well. Because the pathway to understanding cognitive processes is revealed by the analyses of eye movement data, we recorded VL experts' oculomotor behavior while working on the tasks. Clearly this measurement could not be performed in a large classroom survey. Therefore, VL experts were compared to non-experts in Experiment II.

VL experts found as many targets and were as fast as non-experts (H2a, H2b) but differed in the way they found target regions. HMM analysis revealed that experts were able to divide seven of the nine images into the same number of or more areas (hidden states) than novices. Aesthetic interpretation of hidden states suggest that WS does not cover a homogeneous region of “non-attraction” but has meaning and “Gestalt” that goes beyond the gist of the scene and governs clues important for the scene composition. The idea that a sequence of hidden states represents cognitive processes is further emphasized by the visualization of the hidden states of experts with density maps, which revealed semantically meaningful regions on the image. Dissecting the image into these additional regions might help VL experts in understanding the scene composition, best illustrated by specific symbolic objects (angel-figures) or artistic compositions (“Rückenfigur”). These regions might be more salient for VL-experts. Given experts knowledge about image composition and arrangement, experts may be able to “find their way” through the pictures differently. Results hint at fewer hidden states for the non-expert group and a wider spread of fixations across WS compared to experts. This is in accordance with the findings of Koide et al. (2015) stating that experts regarding abstract paintings tend to not only focus their attention on salient regions (for a visual search task, the target and prime regions) but were also able to direct their attention to areas that are disregarded by novices. One possible explanation would stress the role of working memory as the psychological correlate of the underlying cognitive processes during the respective spatiotemporal HMM state. Thus, important information for the image arrangement is processed while fixating a certain region of the image during a given time period (Irwin, 2004).

To assess the precision of expert and non-experts search strategy we assigned hidden states to each pre-defined AOI with the highest fixation overlap. ANOVA revealed a significant interaction of fixation overlap (of hidden state and AOI) with expertise, the image and the pre-defined AOIs. Precision is higher for the expert group on the third prime and target regions (H3). Non-experts show more fixations in previously undefined WS when they look at the third prime and search for the third target region, leading to more fixations outside pre-defined AOIs within the corresponding hidden state, therefore focusing with less precision than the expert group.

Recent evidence has shown that aesthetically appreciated images lead to enhanced perceptual processing (Sarasso et al., 2020). VL experts would have therefore benefitted from the artwork stimuli as it might have improved the engagement and encouraged a deeper commitment into ”analysing” Wagner and Schönau (2016) the image thoroughly. It can also be argued that visual working memory (Bahle et al., 2019; Olivers et al., 2011) of experts form an enhanced representation or mental model of the images.

Hidden Markov models in visual search

Each HMM state in Sample II represents a location on the image participants fixated while inspecting that area during the search. In general, there is a good coincidence between pre-defined AOIs (primes and targets) and the data driven hidden states. WS mostly comprises several distinguishable hidden areas either representing semantically meaningful attractors (e.g. people's faces or control buttons) or distractor regions that had to be excluded (e.g. different angel-figures or different parts of the spider net).

HMMs also allows for an aggregated comparison between fixation sequences. Usually comparing multiple scanpaths between subjects is a complex challenge. E.g. what threshold values should be used to define a starting/landing point for fixations on important image areas? The HMMs based on group level allows for a description of fixation sequences as hidden states on the image, as each fixation point was classified to their most probable hidden region over time.

HMMs have not been extensively used in visual search tasks. However, the method presented here, can be of great use to investigate complex search processes that can go beyond the visual search task presented in this study (e.g. natural scenes, virtual reality and real-world searches) in which classical aggregate statistics may fall short. Subtle differences in viewing behavior can be more clearly defined. The hidden states estimated by the HMM can be interpreted as data-driven AOIs. Instead of defining AOIs by arbitrary thresholds, we can include subject's fixations that lie outside the pre-defined AOIs in WS to be included to any data-driven AOI (hidden state) based on the estimated probability.

The HMM presented here are not exhaustive for eye movement analysis for expertise research and can be expanded upon (e.g. additional variables are conceivable as basis for model formulation such as the length of saccades or individual fixation durations to assess selective attention allocation). Further research is needed in respect to VL skills in the domain of arts, as the literature is dominated by studies in sports and medicine and eye movement differences are heterogeneous across expertise domains (Brams et al., 2019).

Limitations

A few limitations have to be mentioned. First, the images were not tested for low-level saliency (Foulsham, 2019). Our focus was on expert vs. non-expert search strategies in identifying targets from prime regions. Different levels of saliency could interfere with the results presented here (Loftus & Mackworth, 1978). Thus, further studies could vary the number of visual salient features systematically. That way cognitive states prone to bottom-up mechanisms could be differentiated from top-down search strategies guided by expertise and/or working memory.

Another concern for the generalizability of our results is the lack of time constraints combined with a motor coordination problem in identifying the targets. Deviating from earlier studies, this might lead to a varying number of data points per trial. The visual search task was only finished when the participant indicated to have found all three target regions. This led to a wide variance in individual search times untypical for other target search experiments that only last a few seconds (Coutrot et al., 2018). It remains uncertain how many fixations are efficient for HMM for eye movements. Depending on the time constraints and task at hand, different number of fixation points might be necessary. Coutrot et al. (2018) argue that images should contain various regions of interest to capture systematic patterns of exploration behavior. This can be achieved by pre-defined AOI (as in our study) or by using images with different salient areas that draw participants attention.

Lastly, even though the number of participants was above average for eye-tracking studies in expertise research (Gegenfurtner et al., 2011), sample size might still have been too small to find all differences between VL experts and non-experts, especially as effects of VL on visual search behavior seem to be more subtle than proposed by art education research. As the present results are exploratory, further research is needed to confirm these observed differences. There was also a mean age difference of 10 years between each group. However, as the number of fixations was not correlated with age, we would argue that differences in HMMs are not primarily due to age differences. A few seconds of eye movement data used to define fixation sequences was sufficient to model clearly distinguishable hidden states on a group level. This is promising for future use cases with limited sample size and more obvious differences in eye-movement behavior (e.g., in patients with visuospatial neglect (Cox & Aimola Davies, 2020)).

Conclusion and future outlook

This study investigated VL expert and non-expert visual search behavior. The expert group revealed a more detailed search strategy, indicated by a higher number of hidden states and higher precision for looking at the last prime and searching for the last target. Specific image parts, previously not taken into account by pre-defined AOI were outlined in greater clarity among VL experts. Non-experts on the other hand, focused on broader and thus fuzzier image areas during visual search. For the purpose of constructing a VL assessment test battery, selecting more items of intermediate or greater difficulty, including even more realistic and stylistic image compositions, is advised because students displayed a rather skillful ability to perform visual search tasks.

From a methodological view point, the statistical methods used could introduce a new perspective on modeling expertise-related differences in eye movements. Future studies could investigate the link between topological HMM states and “cognitive” hidden states incorporating more variables such as fixation duration or saccadic length into the models. The same idea was followed by van der Lans et al. (2008), who found an association between a 2-state cognitive HMM based on local and global search strategies. Deviating from our art oriented approach in choosing visual stimuli, they used a saliency map based on low-level perceptual features and the scene’s organization to explain their results. Other recent approaches measured oculomotor behavior while switching between hidden cognitive states during a decision task (Chuk et al., 2019). In an educational context, not necessarily restricted to art education, HMMs states could be helpful to describe how much students are “involved” in the given tasks.

5 Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

6 Author Contributions

M.W.G., U.F. and K.R. designed the study. M.T and E.W. selected and prepared the stimuli. M.T. conducted the field work. M.T. and U.F. designed and performed the statistical analysis. E.W. contributed aesthetic theory to the interpretation of statistical results. M.T., M.W.G., K.R. and U.F. prepared the manuscript. All authors reviewed the manuscript.

7 Funding

This study was supported by the German Federal Ministry of Education and Research (grant 01JK1606A).

8 Acknowledgments

We thank students, VL experts and novices for their time participating in this study. This study fulfils part of the requirements for the University of Regensburg doctoral degree program of author M.T.

9 Data Availability Statement

The datasets analyzed during the current study are available from the corresponding author upon request.

Paper 3: Judgment of Visual Abstraction

Assessing Heterogeneity in Students' Visual Judgment: Model-Based Partitioning of Image Rankings

Miles Tallon, Mark W. Greenlee, Ernst Wagner, Katrin Rakoczy, Wolfgang Wiedermann, Ulrich Frick

Author-produced version of this article under peer review for publication in Frontiers in Psychology: Methods and Applications in: Perception Science.

The published peer-reviewed version of this article is available in Frontiers in Psychology: https://doi.org/10.3389/fpsyg.2022.881558

Abstract

Differences in the ability of students to judge images can be assessed by analyzing the individual preference order (ranking) of images. To gain insights into potential heterogeneity in judgment of visual abstraction among students, we combine Bradley-Terry preference modeling and model-based recursive partitioning. A sample of 1020 high-school students ranked five sets of images, three of which with respect to their level of visual abstraction. Additionally, 24 art experts and 25 novices were given the same task, while their eye movements were recorded. Results show that time spent on the task, the students' age, and self-reported interest in visual puzzles had significant influence on rankings. Fixation time of experts and novices revealed that both groups paid more attention to ambiguous images. The presented approach makes the underlying latent scale of visual judgments quantifiable.

Keywords: visual abstraction, assessment, Bradley Terry model, model-based partitioning, ranking, art education, visual literacy

1 Judgment of Visual Abstraction

Every art, whether figurative or not, is a form of abstraction (Gortais, 2003; Witkin, 1983). However, the measurement of the perceived level of visual abstraction in artworks remains challenging. Based on psychological models of aesthetic judgment, the judgment of visual abstraction would be part of the continuous affective evaluation proposed by the model of aesthetic experience (Leder et al., 2004; Leder & Nadal, 2014) comprising the early processing stage of perceptual analysis and explicit classification of content and style of visual art. In an experiments with pairs of modern art paintings that had to be evaluated for similarity, Augustin et al. (2008) found that the content (motif) of paintings is processed at presentations as early as 10ms, while processing the artists' style can be observed at presentations of 50ms and onwards. This initial stage can also include the appraisal of visual abstraction in images of artwork.

Previous studies explored the preference judgment of abstract art measured by Likert-scale ratings and revealed a preference for the artists' original compositions (Furnham & Rao, 2002; McManus et al., 1993). Art experts seem to prefer abstract over realistic representations (Gartus et al., 2020; Hekkert & van Wieringen, 1996; Silvia, 2006), whereas novices prefer representational paintings (Pihko et al., 2011). Layperson often express a higher sense of meaning for representational art than abstract art (Schepman et al., 2015; Schepman & Rodway, 2021). Studies in computer science used algorithmic approaches to construct and identify the level of visual abstraction in images and artworks; e.g. abstraction in portrait sketches (Berger et al., 2013; Muhammad et al., 2018). Other studies have demonstrated how concrete representations of icons are more beneficial for visual programming than abstract icons (García et al., 1994). Visual abstraction can also foster visual communication skills (Fan et al., 2020). E.g., pictograms (Tijus et al., 2007) as a form of visual communication, can be based on visual similarity (i.e., reproducing the visual characteristics of an object), semantic association (e.g., an image of a clock to convey the concept of "time"), and arbitrary convention (connection established through verbal reinforcement). Nakamura and Zeng-Treitler (2012) conducted a taxonomical analysis of over 800 health-related pictograms and concluded that semantic association is the most used strategy and the only effective way for pictograms to represent abstract concepts such as *love* or *pain*.

A study that specifically tried to measure the perceived level of visual abstraction used visual analog scales to rate artworks as "abstract" and found contrast effects due to sequential presentation of high vs. low abstract paintings on the judgment (Specht, 2007). Efforts quantifying visual abstraction in artworks were also done by Chatterjee et al. (2010): the

Assessment of Art Attributes instrument (AAA) includes “abstraction” as a conceptual-representational attribute. The level of abstraction is measured through a Likert-scale rating and training slides with example images as anchors. Another assessment tool, the Rating Instrument for Two-Dimensional Pictorial Works (RizbA) (Schoch & Ostermann, 2020), consists of 26 six-point Likert-scale items, including two questions regarding the manner of concrete and abstract representation.

However, aesthetic judgments are not only influenced by the properties of the items being judged but it is influenced by additional factors such as expertise and personal experience (Child, 1965; Hayn-Leichsenring et al., 2020; Jacobsen, 2004; McCormack et al., 2021; Nodine et al., 1993). Chamorro-Premuzic and Furnham (2004) showed how university students with higher interests in art tend to score higher on art judgment scores and that these judgments were significantly related to both personality and intelligence. Identifying critical variables that influence students’ judgments of visual abstraction may represent an important milestone for empirical art education research.

Assessment of Latent Image Characteristics for Ranking Tasks

When underlying image features are latent (e.g., the extent to which a given image is abstract) metric scales may fall short when asked to judge these items, by, for example, assigning them a number from 1 to 10. Typical disadvantages of the use of such absolute measures may include anchor effects (Furnham & Boo, 2011) and end-aversion bias (Streiner & Norman, 2008) among others (Choi & Pak, 2005). It is oftentimes easier to compare items to each other, e.g., in a series of paired comparison (PC) tasks. Such comparative measures can be analyzed with Bradley Terry (BT) models (Bradley & Terry, 1952), also referred to as Bradley-Terry-Luce models. BT models are a popular method to uncover a latent preference scale of objects/items from paired comparison data (Cattelan, 2012). For example, BT models are frequently used to determine the best sport teams (Cattelan et al., 2013), to analyze consumer-specific preferences (Dittrich et al., 2000), or to elicit perceived harm of psychotropic substances (Wiedermann et al., 2014). When multiple objects (images) are compared simultaneously, ranking tasks (e.g., ranking images according to their level of abstraction) constitute valuable alternatives to PCs. Ranking data can then be transformed into derived PC patterns (Francis et al., 2010).

An innovative approach is used that combines BT models with model-based recursive partitioning (trees) to detect preference heterogeneity in subgroups (Wiedermann et al., 2021). BT models can be used for (art) educational assessment tasks, in which students are instructed to rank images based on given criteria. From a methodological perspective the use of BT models in combination with recursive partitioning holds great potential when applied to art education assessment: conventional statistical analysis of interaction effects may fall short when tasked to address the complex moderation processes of visual judgments. This method enables researchers to differentiate between the effects of student characteristics and learning interventions on latent preference rankings more closely.

Current Study

This study is part of a larger research project on the assessment of Visual Literacy (VL) and how VL can be fostered in art education (Frick, Rakoczy, Tallon, Weiß, & Wagner, 2020). VL, a core competency in art education, comprises the ability to evaluate artwork with respect to aesthetic value. The Common European Framework of Reference for Visual Literacy (CEFR-VL; Wagner & Schönau, 2016) defines *judging* (or evaluating) images as the ability to formulate a justified statement or estimation about images and artistic creations.

The aim of the present study is to investigate students' ability, on the one hand, as well as that of experts and novices, on the other hand, to judge images based on the level of perceived visual abstraction, while placing a focus on the discovery of individual factors that influence these judgments. The ability of visual abstraction is assumed to be as important for visualizations as it is for analytic thinking (Punzalan, 2018; Viola et al., 2020). The results may help to determine essential variables that impact the judgment of abstraction and in return they might be able to help teachers detect and promote students' development of artistic skills.

The potential heterogeneity in perceived visual abstraction was evaluated in two samples: A sample of high school students and a further sample comprising art experts (art educators, artists, designers) and novices (art laypersons). In the student sample, self-reported visual skills and demographic variables are used to detect potential differences in students' performance to rank different sets of images based on level of abstraction. In the experts and novices sample eye movements were recorded during the image ranking task. Eye movement indicators are used to analyze the distribution of attention (Brams et al., 2019; Jarodzka et al., 2017). Eye tracking, especially as an exploratory tool, can enhance the multidisciplinary field of VL research, as it visualizes cognitive processes involved in visual problem solving and art perception (Brumberger, 2021). Visualizing the solution process with VL-expert's and novices' eye-movements can be used to uncover cognitive processes that differ between the expert and novice groups and may further point out difficult or ambiguous image sets.

This study addresses the following research questions: What effects do self-reported visual skills and student characteristics have on the order of images ranked according to visual abstraction? Do VL-experts and novices differ in their ranking patterns and solution strategies?

2 Methods

Subjects and Stimuli

Sample I comprised 1020 students of which 987 worked on the ranking tasks and filled out the questionnaire. A total of 52 classes (9th to 13th grade) from 29 schools in Germany took part in the study. Two classes did not receive the questionnaire and one class could not be offered the ranking task due to technical difficulties. To control for potentially nested effects of classrooms, intraclass correlation coefficients (ICCs) for correct rankings were calculated on each image set. Due to low values (ICCs range from 0.01 to 0.03, for calculations see Chakraborty & Sen, 2016), no multi-level adjustments were necessary. Overall, 52% of participants were female, the average age was 15.34 years (*SD* = 2.96). Schools were recruited in the federal states of Hessen, North-Rhine Westphalia, Schleswig-Holstein, and Rhineland Palatinate via leaflets, letters and recommendations. Data collection was conducted in classrooms with up to 30 students (*M*= 20.8, *SD*= 5.10). The image ranking task was part of a VL assessment test battery, including demographic questions, art grade, and the following questions regarding artistic ability and self-perceived art skills (S1-S5):

- If you had to rank all of your classmates according to their abilities in the subject of art, where would you rank yourself? (S1; scored 1 [as one of the worst] to 5 [as one of the best])
- How good are you at art in general? (S2; scored 1 [very bad] to 5 [very good]))
- How good are you in theoretical content (art theory; e.g. interpreting pictures, understanding art history)? (S3; scored 1 [very bad] to 5 [very good]))
- How good are you in practical activities in art class (e.g. painting, drawing, drafting, and designing)? (S4; scored 1 [very bad] to 5 [very good])
- Compared to your skills in other school subjects: How well do you rate your art skills? (S5; scored 1 [much worse] to 5 [much better])

Additionally the following self-reported visual skills were rated on a scale from 1 (strongly disagree) to 4 (strongly agree): Photographic memory (PM): “I have a 'photographic memory'”;

Spatial orientation (SO): “When I see a photograph of a geometric object, I can imagine what it looks like from behind”); Long-term memory (LM): “I can remember small details in pictures”; Imagination (IM): “I can easily picture things mentally”; and Interest in visual puzzles (IP): “I like to solve picture puzzles”.

Sample II comprised 51 participants of which 49 participants had qualitatively sufficient eye-tracking data to be included for further analyses. Experts and novices were screened based on their experience and interest or profession in the visual arts. The expert group (n = 24) consisted of photographers, artists, designers, and art students. The novice group (n = 25) consisted of students and adults from various educational institutions who were not associated with academic or professional work in the visual arts. The mean age of participants were M = 29.08 years (SD = 12.55). The participants in sample II were assessed individually in seminar or laboratory rooms (e.g., at the Academy of Fine Arts in Munich).

Ranking Task

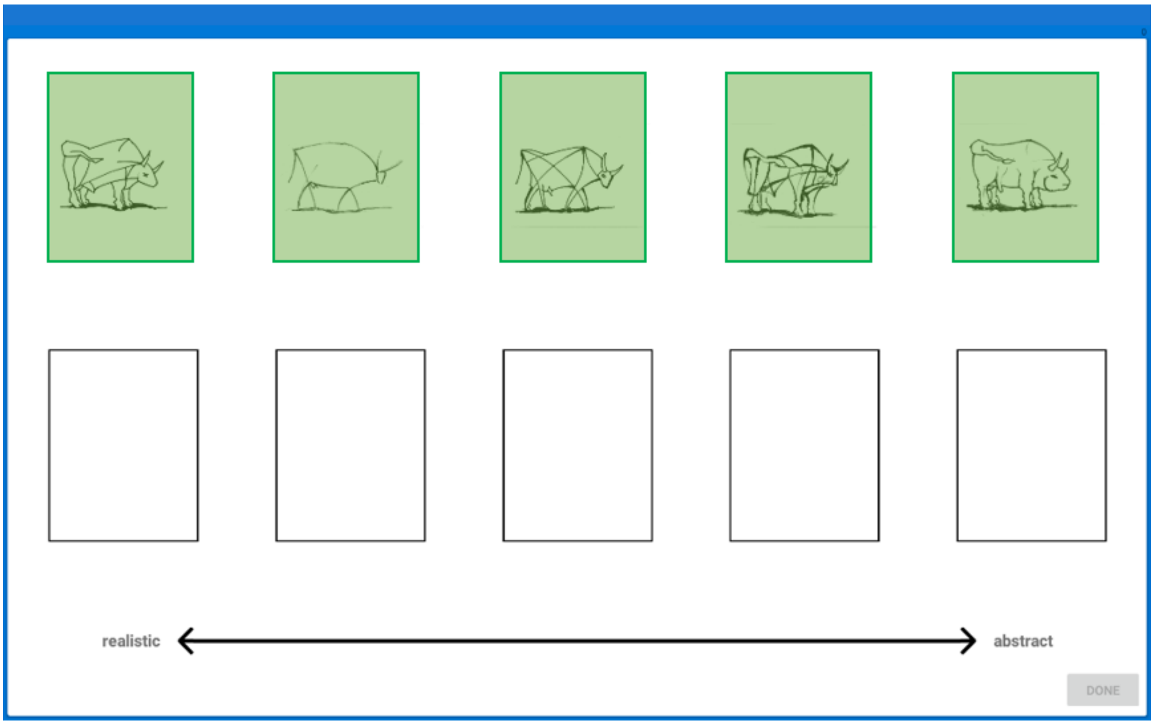
We used images with varying level of visual abstraction, i.e. image sets that represent the gradual process of transforming figurative artwork to non-figurative artwork (Viola et al., 2020). As every work of art uses some level of abstraction, many artworks could be investigated. Therefore images were curated (or created) by visual arts professionals from the board of the European Network for Visual Literacy (ENViL). Image sets were chosen based on the likelihood of being discussed in art class, representing a varying degree of abstraction. Overall, five ranking tasks were presented on Android tablets with 10.1 inch screen size (Andrews et al., 2018). Subjects ranked 5 images, resulting in a total of $\binom{5}{2} = 10$ paired comparisons for each set of images (with a total of 5! = 120 possible combinations; see Table 1). All participants were presented with the same initial ordering of images and were instructed to rank each image according to two characteristics presented below each image set. The image sets included:

1. geometric figures
2. dogs
3. bull images, inspired by Pablo Picasso’s *Bull* lithographs (MacTaggart, 2021)
4. Mondrian trees
5. salt packages (only presented in sample I)

Images had to be ranked according to the following image characteristics: starting with an example item ("geometric figures"), from round to edgy, the items "dogs", "bull images" and "Mondrian trees" had to be ranked by level of abstraction; from most realistic to most abstract. Additionally, as a control condition, perceived expensiveness; from cheap to expensive ("salt packages") was assessed. In contrast to the evaluation of image abstraction, rankings based on unknown prices should stand out as visible outliers compared to the other rankings. Participants used the touchscreen to select and drop each image into empty slots below (see Fig. 1). The image rankings are then analyzed to gain insights into the possible effects of the participant characteristics on the perceived judgment of abstraction.

Figure 1

Ranking Item Bull Images



Note. Each bull picture above needs to be placed into an empty slot below to form a ranking from most realistic (left) to most abstract (right). Areas of Interest (AOIs) were not visible by subjects.

In sample I school classes were offered a lump sum of 100€ as collective compensation. In sample II student participants each received 20€ as compensation. Participants from the expert group, who were generally interested in the subject of visual literacy and eye tracking, took part without further incentive. All participants and their legal representatives respectively gave written consent before participating in this study. The study was conducted according to the guidelines for human research outlined by the Declaration of Helsinki and was approved by the Ethics Committee of Research of the Leibniz Institute for Research and Information in Education, Frankfurt am Main (DIPF, 01JK1606A).

Eye Tracking

Each participant in sample II wore eye-tracking glasses (SMI ETG 2w Analysis Pro) during task performance. Eye movements were recorded at 60Hz. A 3-point calibration was performed on the tablet for each participant. All participants had normal or corrected to normal eyesight. Fixations were mapped onto corresponding reference images using SMI fixation-by-fixation semantic gaze mapping (Vansteenkiste et al., 2015). Areas of Interest (AOIs) were drawn on each image to assess fixation time and number of fixations spend on each image. Eye-movement events were determined by the SMI velocity-based algorithm (Engbert et al., 2016). Eye-tracking data, i.e., number of fixations, fixation duration and heatmaps were analyzed with SMI BeGaze version 3.7. Heatmaps are used as exploratory tools to investigate eye movements (Bojko, 2009) supplementing the BT models.

Data Analytic Strategy

We used Bradley Terry (BT) models as the basis for recursive partitioning. The BT model is a probability model that can be used to predict the outcome of paired comparisons and to obtain (cardinal) preferences values for all items (images) on a latent scale (Bradley & Terry, 1952). Here, “preference” refers to the judgment of image characteristics (e.g. abstractness) by each participant. The probability of preferring item *j* over item *k* can be described as

$$p_{j>k} = \frac{\pi_j}{\pi_j + \pi_k} \quad (1)$$

with π representing the “worth” of the item, quantifying the position of the item on a standardized latent scale from 0 to 1. The worth parameters (π) indicate how likely an item is selected in a paired comparison. BT models can be fitted as loglinear Bradley Terry models (LLBT) (Sinclair, 1982; Dittrich et al., 1998). In the basic LLBT, the linear predictor ŋ is given by

$$ŋ_{y_{jk}} = \ln m(y_{jk}) = \mu_{jk} + y_{jk}(\lambda_j - \lambda_k) \quad (2)$$

where m denotes the expected frequency of PC decisions, μ_{jk} is a nuisance parameter for the comparison jk which fixes the marginal distribution to n_{jk} and y_{jk} are indicator variables with value 1, if object j is preferred to k and value -1, if object k is preferred to j. The λ parameters can be transformed into worth parameters by the equation

$$\pi_j = \exp(2\lambda_j)/\sum_k \exp(2\lambda_k). \tag{3}$$

As the ranking responses of a subject are considered simultaneously a pattern approach is used. The response pattern is defined as $\boldsymbol{y} = (y_{12}, y_{13}, \ldots, y_{jk}, \ldots, y_{J-1,J})$. The expected frequency for a sequence of preferences $\boldsymbol{y}$, formulated as a loglinear model, is given as

$$m(\boldsymbol{y}) = m(y_{12}, \ldots, y_{J-1,J}) = np(\boldsymbol{y}) \tag{4}$$

where n is the total number of respondents and $p(\boldsymbol{y})$denotes the probability to observe the response pattern $\boldsymbol{y}$.

To gain PC patterns of rankings, rankings are converted into a series of paired comparison decisions (Dittrich et al. 1998). Note that in the case of forced rankings (i.e., no mid-ranks), ties do not occur by definition. Rankings are transformed into a series of paired comparisons of which intransitive patterns (e.g. 1 > 2 and 2 > 3, but 3 >1) cannot occur and as such are reduced to J! possible combinations (Dittrich et al., 2002). Model parameters are estimated using a log link and a Poisson-distributed error component. Table 1 shows the design structure of the LLBT model.

Table 1

Design Structure of the Loglinear BT Pattern Model for Rankings Obtained from J = 5 Images

	Paired Comparison (PC) Patterns											Model Parameters					
Rankings	y_{12}	y_{13}	y_{14}	y_{15}	y_{23}	y_{24}	y_{25}	y_{34}	y_{35}	y_{45}	counts	intercept	x_1	x_2	x_3	x_4	x_5
a b c d e	1	1	1	1	1	1	1	1	1	1	n_1	1	4	2	0	-2	-4
b a c d e	–1	1	1	1	1	1	1	1	1	1	n_2	1	2	4	0	-2	-4
c a b d e	–1	–1	1	1	–1	1	1	1	1	1	n_3	1	2	0	4	-2	-4
…	…	…	…	…	…	…	…	…	…	…	...	…	…	…	…	…	…
c e d b a	–1	–1	–1	–1	–1	–1	–1	1	1	–1	n_{118}	1	–4	–2	4	0	2
d e c b a	–1	–1	–1	–1	–1	–1	–1	–1	–1	1	n_{119}	1	–4	–2	0	4	2
e d c b a	–1	–1	–1	–1	–1	–1	–1	–1	–1	–1	n_{120}	1	–4	–2	0	2	4

Note. Rankings are transformed into paired comparison (PC) patterns; the y's represent obtained PCs (y_{jk} = 1 if $j > k$ and y_{jk} = –1 if $k > j$), each possible combination of J! is then counted as observed frequencies in column "counts", and *x*'s are auxiliary variables used to estimate model parameters indicating how often *j* was preferred minus how often *j* was not preferred.

To incorporate subject covariates in BT models we used model-based recursive partitioning (MOB; Zeileis et al., 2008) to identify groups of subjects that differ in their preference rankings. The covariate space is recursively divided (partitioned) into sub-groups of subjects with varying image rankings to form a tree-structured division (Strobl et al., 2011). Each terminal node of the tree structure consists of a separate LLBT model with partition-specific model parameters. Wiedermann et al. (2021) extended the MOB BT framework to distinguish between focal independent variables (e.g., expertise status) and covariates used for recursive partitioning. The MOB LLBT model for g = 1, …, G subgroups can be written as

$$\log\left[m(y_{jk})_{(g)}\right] = \mu_{(g)} + \lambda_{s(g)} + y_{jk|s(g)}\left(\lambda_{j(g)} + \lambda_{js(g)} - \lambda_{k(g)} - \lambda_{ks(g)}\right) \qquad (5)$$

where the intercept $\mu_{(g)}$ and the main effect $\lambda_{s(g)}$ constitute normalizing constants in subgroup g, $y_{jk|s(g)}$ gives the paired comparison decision in group s and partition g (with $y_{jk|s(g)}$ = 1 if $j \succ k$ and $y_{jk|s(g)}$ = –1 if $k \succ j$), $\lambda_{j(g)}$ and $\lambda_{k(g)}$ denote the partition-specific object parameters for the reference group, and $\lambda_{js(g)}$ and $\lambda_{ks(g)}$ are the partition-specific effects capturing potential group differences (c.f. Wiedermann et al., 2021).

Covariates are included to assess the additive impact of subjects' characteristics on the perceived worth of image features. Students in sample I include the following covariates: The time spent on each image set ("Game Time"), gender, age, art grade, and the questions regarding artistic ability and self-perceived art skills. Sample II covariates included age, gender, time spent on each image set, and eye-tracking variables fixation time (time spent fixating image AOIs) and fixation counts (fixations lying inside image AOIs). VL expertise status (expert vs. novice) served as a focal independent variable.

Statistical analysis and model formulation were conducted with the R-package "prefmod" (Hatzinger & Dittrich, 2012), partitioning was accomplished with the R-package "partykit" (Hothorn & Zeileis, 2015). To overcome the risk of spurious tree structures a minimum node size of 40 was chosen for Sample I and a minimum of 4 participants for Sample II to reduce model complexity. To avoid overfitting, a post-pruning strategy based on the Akaike Information Criterion (AIC) was used to prune splits (i.e., bifurcations) that do not improve model fit (Zeileis et al., 2008). Non-parametric bootstrapping (using 1000 resamples) was used to evaluate the stability of LLBT trees (Philipp et al., 2018). Here, we focused on selection probabilities and average cut-off (splitting) values of the pre-defined covariates. For a stable LLBT tree, selection probabilities of the initially selected covariates are expected to be close to one and average splitting values are expected to be close to the estimates obtained in the initial LLBT tree.

3 Results

Student Sample I

Table 2 shows the descriptive statistics for self-reported variables and time spent on each image set for sample I. Depending on the image set, different variables had significant impact on the preference rankings.

Table 2

Descriptive Statistics of Variables in Sample I (N= 987 Students)

Variable	Mean (SD)	
Age	15.35 (2.96)	
S1	3.63 (0.97)	
S2	3.70 (0.89)	
S3	3.33 (0.95)	
S4	3.70 (1.08)	
S5	3.26 (1.16)	
PM	2.57 (0.88)	
SO	3.20 (0.76)	
LM	2.70 (0.8)	
IM	2.05 (0.93)	
IP	2.74 (0.91)	
Art grade	1.96 (0.84)	
Mean time on…		Percentage of correct* ranking
…Geometric figures	13.28 (5.45)	96 %
…Dogs	23.01 (10.26)	42 %
…Bull images	24.33 (12.71)	29 %
…Mondrian trees	18.16 (9.05)	36 %
…Salt packages	27.46 (14.49)	04 %

Note. S1-S5 = self-perceived art skills, PM = Photographic Memory, SO = Spatial Orientation, LM = Long-term Memory, IM = Imagination, IP = Interest in visual Puzzles. *intended ranking: a>b>c>d>e

Table 3 shows the worth parameters for the LLBT tree terminal node in each image set, including significant splitting covariates for sample I. Worth parameters (π) range from 0 to 1, and sum up to 1 for each node. For most image sets, exception being the "salt packages" and the "bull images", worth parameters decline and form a slope from highest worth to lowest worth according to the intended solution for each image set.

Table 3

Worth Parameters in each Terminal Node from Sample I

Sample I – Students (n=987)							
Image set	Term. node	Worth parameters (π) for each image					Splitting covariates
		a	b	c	d	e	
Geometric figures	n=634	0.933	0.061	0.005	4.10E-04	2.00E-05	Age <= 15
	n=259	0.921	0.069	0.007	9.00E-04	6.30E-05	Age >15, Time <=15 sec
	n=94	0.593	0.228	0.106	0.053	0.018	Age > 15, Time > 15 sec
Dogs	n=182	0.415	0.241	0.143	0.120	0.081	Time<= 20 sec, IP <= 2
	n=312	0.318	0.237	0.184	0.144	0.117	Time <= 20 sec, IP > 2
	n=46	0.280	0.233	0.230	0.158	0.099	Time >20 sec, IP <= 1
	n=447	0.403	0.223	0.184	0.116	0.074	Time >20 sec, IP > 1
Bull images	n=76	0.403	0.226	0.157	0.134	0.080	Time <= 12 sec
	n=911	0.585	0.186	0.091	0.099	0.038	Time > 12 sec
Mondrian trees	n=59	0.577	0.157	0.135	0.073	0.058	Time<13 sec, Age<= 14
	n=117	0.509	0.176	0.182	0.081	0.053	Time<13 sec, Age> 14, LM <= 2
	n=158	0.831	0.077	0.074	0.013	0.004	Time< 13 sec, Age> 14, LM > 2
	n=654	0.624	0.144	0.136	0.052	0.043	Time > 13 sec
Salt-packages	n=450	0.274	0.325	0.144	0.127	0.130	male
	n=495	0.325	0.347	0.113	0.109	0.107	female

Note. IP = Interest in visual puzzles, LM= "I can remember small details in pictures" from 1 (strongly disagree) to 4 (strongly agree)

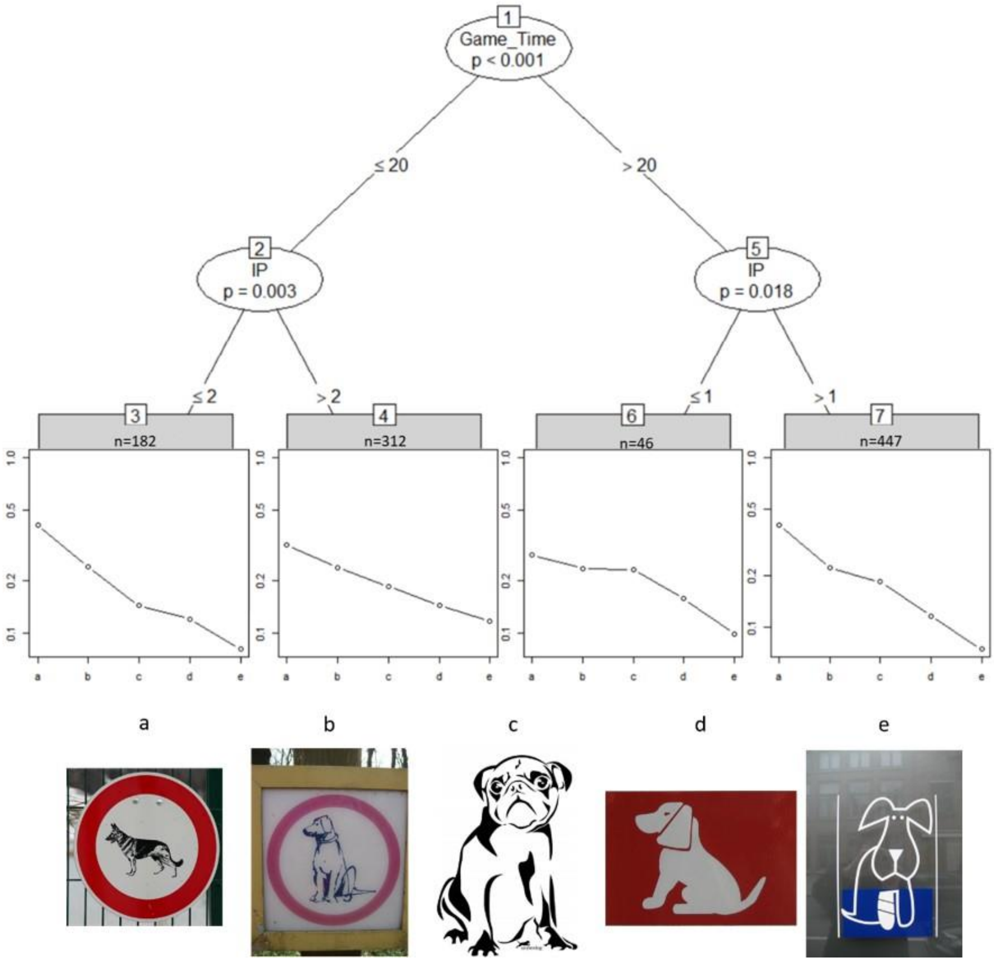
Note that at first glance, certain image sets with worth parameters close to zero would indicate no preference for any of these images. However, this is due to the continuous transformation of the BT model parameters (λ) into a worth parameter (π) on a scale from 0 and 1. For example, for the image set "geometric figures", each image in the first terminal node (n=634 students) is about 12-20 times more likely to be judged to be more "round" compared to the preceding image in the order “a then b then c then d then e”. Image c, (with π=0.005) is about 82% more likely to be chosen before image d (π=0.00041) from participants in the first terminal node.

Overall, the time spent on each set and the participants’ age had the largest impact on the perceived image features. In general, faster and older student groups tend to form the steepest decline in worth parameters between each image, i.e., image preferences between each image are more clearly separated, indicating no problems in ranking the images according to the intended features. Interestingly, two self-reported visual skills “Interest in visual puzzles” (IP) and “long-term memory” (LM) were important for the judgment of abstraction (i.e., ranking images from realistic to abstract) on item set “dogs” and item set “Mondrian trees”. Here, subgroups with higher scores tended to show steeper decline in worth parameters.

Figure 2 shows the partitioning tree for the dog images. The worth parameter is presented on a log-scale. The student sample is split between fast and slow student groups (about 50%) with one group spending less than 20 seconds on the image set (Game_Time < 20) and the other group going above 20 seconds. The gap in perceived abstraction level between dog image b and c is more difficult to differentiate for students in node 6 and 7, i.e. slower student groups have more problems differentiating between the two. However, an interest in visual puzzles (IP>1) helps slower students (45%) realize how image c is less realistic than image b.

Figure 2

Partitioned Paired Comparison Tree for the Ranking Task "Dogs" in Sample I

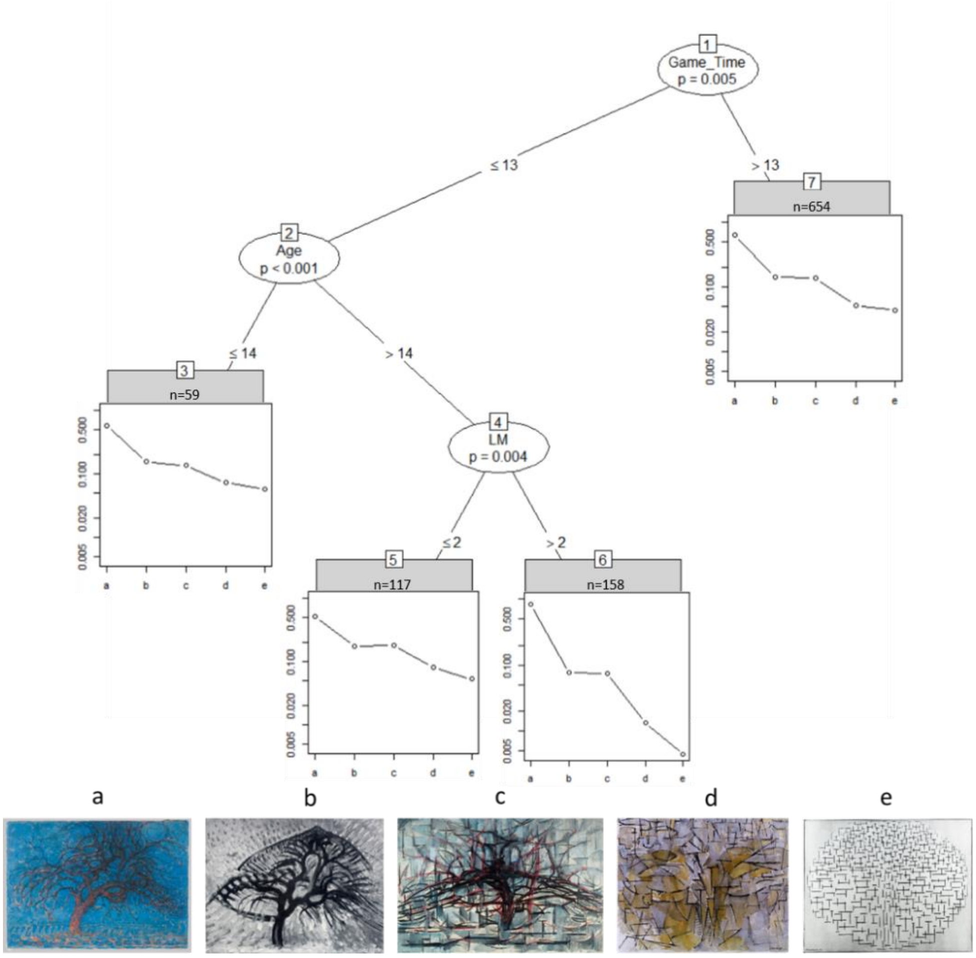


Note. Game_Time= Time spent on image set in seconds, IP="Interest in visual puzzles". Fast students (<20 seconds) show greater differentiating skill between dog image b and c than slow students (>20 seconds). Self-reported IP greater than 1 can increase the perceived differences between dog image b and c (node 7), even in slower student groups.

Figure 3 shows how time spent on the task significantly affects the way students in sample I ranked the tree images from realistic (left) to abstract (right). Most students took longer than 13 seconds to rank the images (n = 654 in node 7) and ranked images b and c close to each other. Faster students under the age of 15 also ranked the tree images according to their proposed level of abstraction (node 3). Older students with self-reported low long term visual memory skill (LM; disagreeing to the statement "I can remember small details in pictures") rate image c to be more realistic than image b (node 5). When these students were agreeing or strongly agreeing to that statement instead (node 6) they rated the first image (a) to be nearly 11 times more realistic than the second image (b) and the last image (e) to be about 3 times more abstract than the fourth image (d).

Figure 3

Partitioned Paired Comparison Tree for the Ranking Task „Mondrian Trees" in Sample I



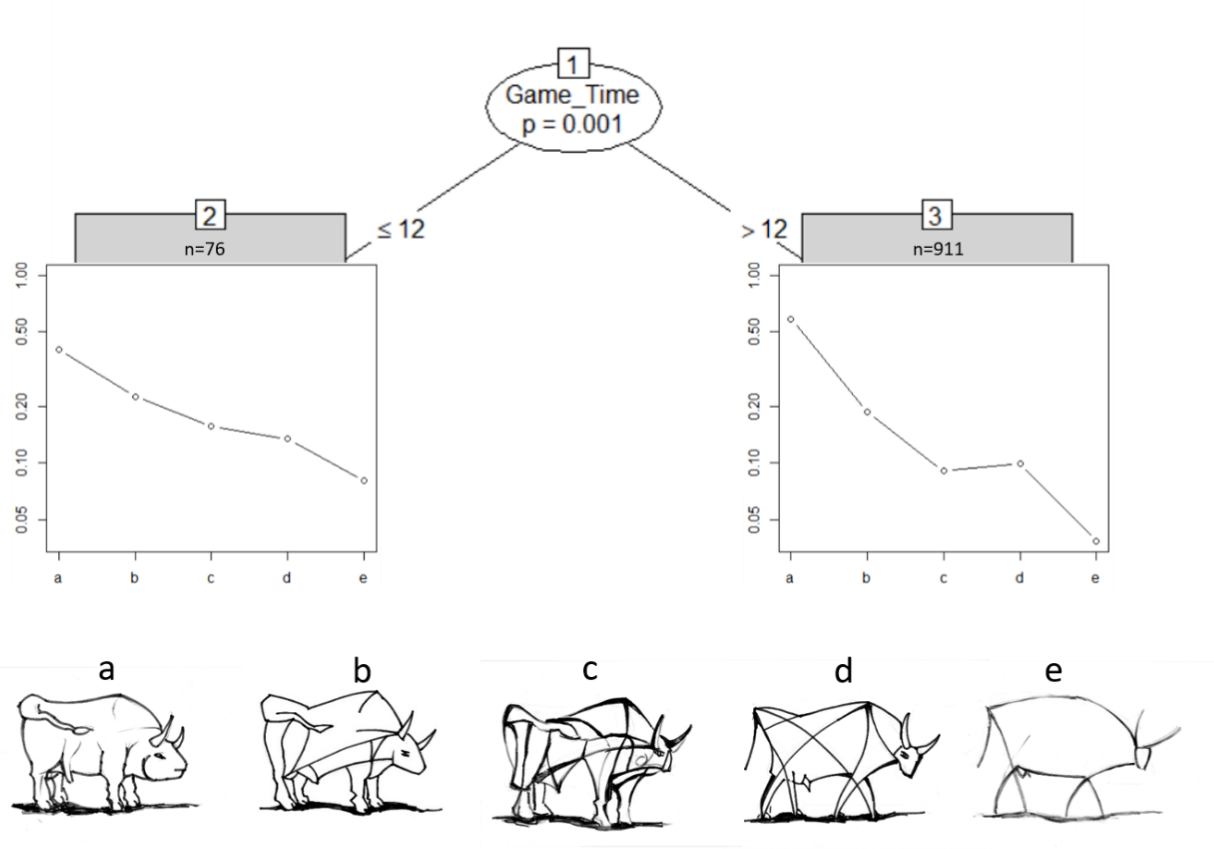
Note. Game_Time = Time spent on image set in seconds, LM= "I can remember small details in pictures"

Figure 4 shows the partitioned tree for the "bull images" set for sample I. Surprisingly, most students (92%) took longer than 12 seconds and rated image d to be more realistic than image c. The "bull image" set is the only image set with a clear deviation from the intended solution.

Figure 5 shows how the cost of salt packages is clearly split between images "a, b" versus" c, d, and e". There is also a significant difference in gender: male students rank the salt image "b" to be less expensive than the female students.

Figure 4

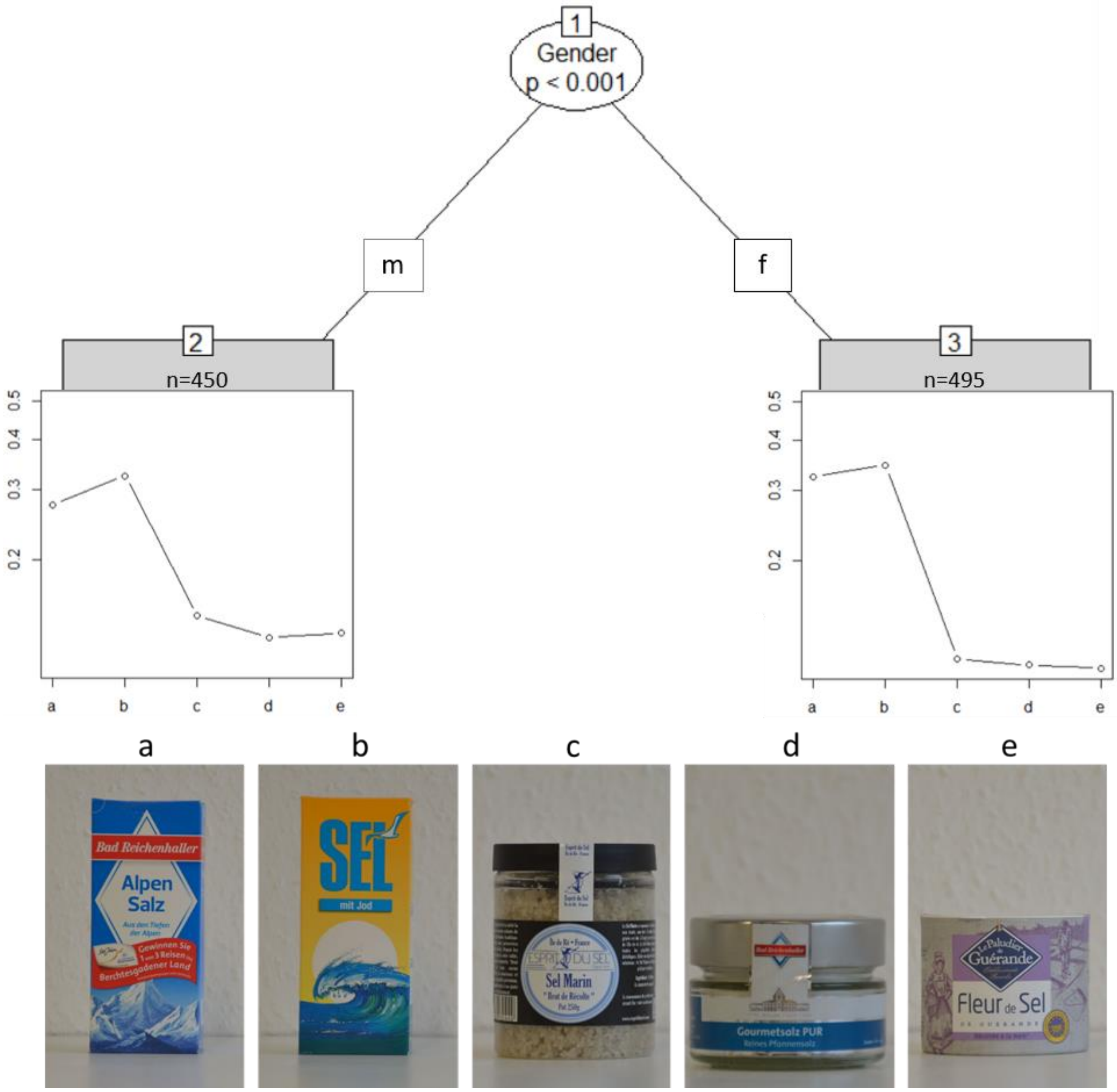
Partitioned Paired Comparison Tree for the Ranking Task „Bull Images“ in Sample I



Note. Game_Time = Time spent on image set in seconds.

Figure 5

Partitioned Paired Comparison Tree for the Ranking Task „Salt Packages“ (Sample I)



Note. m=male, f=female.

Robustness

Stability checks were performed with a bootstrapping procedure, using 1000 bootstrap samples. Table 4 shows the probability of splits based on each covariate in sample I and sample II. In sample I, usually, the time spent on each image set was a common splitting variable, oftentimes splitting the decision tree on each image set except for the “Geometric figures”. Students’ age had significant influence on the stimuli “bull images” and the “Mondrian trees”.

Table 4

Selection Probabilities of Splits for each Variable on each Image Set for Bootstrapping Procedure on Sample I and Sample II

Probability to split tree					
Variable	Geometric figures	Dogs	Bull images	Mondrian trees	Salt packages
Sample I (n= 987 students)					
Age	0.14	0.49	**0.60**	**0.65**	0.45
Gender	0.08	0.42	**0.79**	0.52	**0.92**
Game Time	0.32	**0.98**	**0.98**	**0.92**	**0.82**
Art grade	0.16	0.33	0.35	0.39	0.31
S1	0.11	0.25	0.41	0.35	0.30
S2	0.08	0.44	0.52	0.52	0.43
S3	0.28	0.30	0.33	0.30	0.22
S4	0.12	0.27	0.45	0.55	0.42
S5	0.02	0.34	0.29	0.48	0.29
PM	0.29	0.48	0.42	0.44	0.34
SO	0.02	0.40	0.39	0.56	0.45
LM	0.27	0.34	0.42	0.44	0.33
IM	0.11	0.54	0.53	**0.61**	**0.62**
IP	0.18	**0.82**	**0.62**	**0.66**	0.45
Sample II (n= 49 VL-experts and novices)					
Age	0.00	0.20	0.40	0.17	-
Gender	0.00	0.01	0.00	0.00	-
Game Time	0.00	0.48	0.03	0.06	-
Fix. duration a	0.00	0.19	0.07	0.05	-
Fix. duration b	0.00	0.03	0.00	0.00	-
Fix. duration c	0.00	0.04	0.00	0.05	-
Fix. duration d	0.00	0.02	0.01	0.15	-
Fix. duration e	0.00	0.06	0.01	0.00	-
Fix. count a	0.00	0.39	0.01	0.01	-
Fix. count b	0.00	0.28	0.01	0.00	-
Fix. count c	0.00	0.08	0.00	0.03	-
Fix. count d	0.00	0.09	0.05	0.01	-
Fix. count e	0.00	0.05	0.02	0.00	-

Note. Probabilities of splits >0.60 are marked in **bold**. S1-S5 = self-perceived art skills, PM = Photographic memory, SO = Spatial orientation, LM = Long-term memory, IM = Imagination, IP = Interest in visual puzzles, a= most realistic image to e= most abstract image.

The stability checks indicate that the results from the empirical sample I are comparable: multiple splits on the same decision tree are frequently caused by the time spent on each image set. Questionnaire items S1-S5 on self-reported artistic ability do not seem to trigger splits very often. A few exceptions are noticeable: for the “Mondrian trees” the self-reported ability to imagine (IM) was observed more often to cause a split (M=0.61) in comparison to the long-term working memory (LM) variable (M=0.44) that is reported in the empirical sample. IM was also nearly equally often used to split the tree of the “Salt packages” stimuli. Additionally, interest in visual puzzles (IP) was also found to split variables on the “bull images” and “Mondrian trees” (>60%), therefore might being underrepresented by the empirical sample. Bootstrapping results for the expert and novices in sample II indicate low splitting probabilities (<15%) for the eye-tracking variables. An exception being the “dogs” image set with fixations on the most realistic image splitting the tree in about 40 percent of the time. Lastly, the time spent on the dog images was significant in about 50% of the cases.

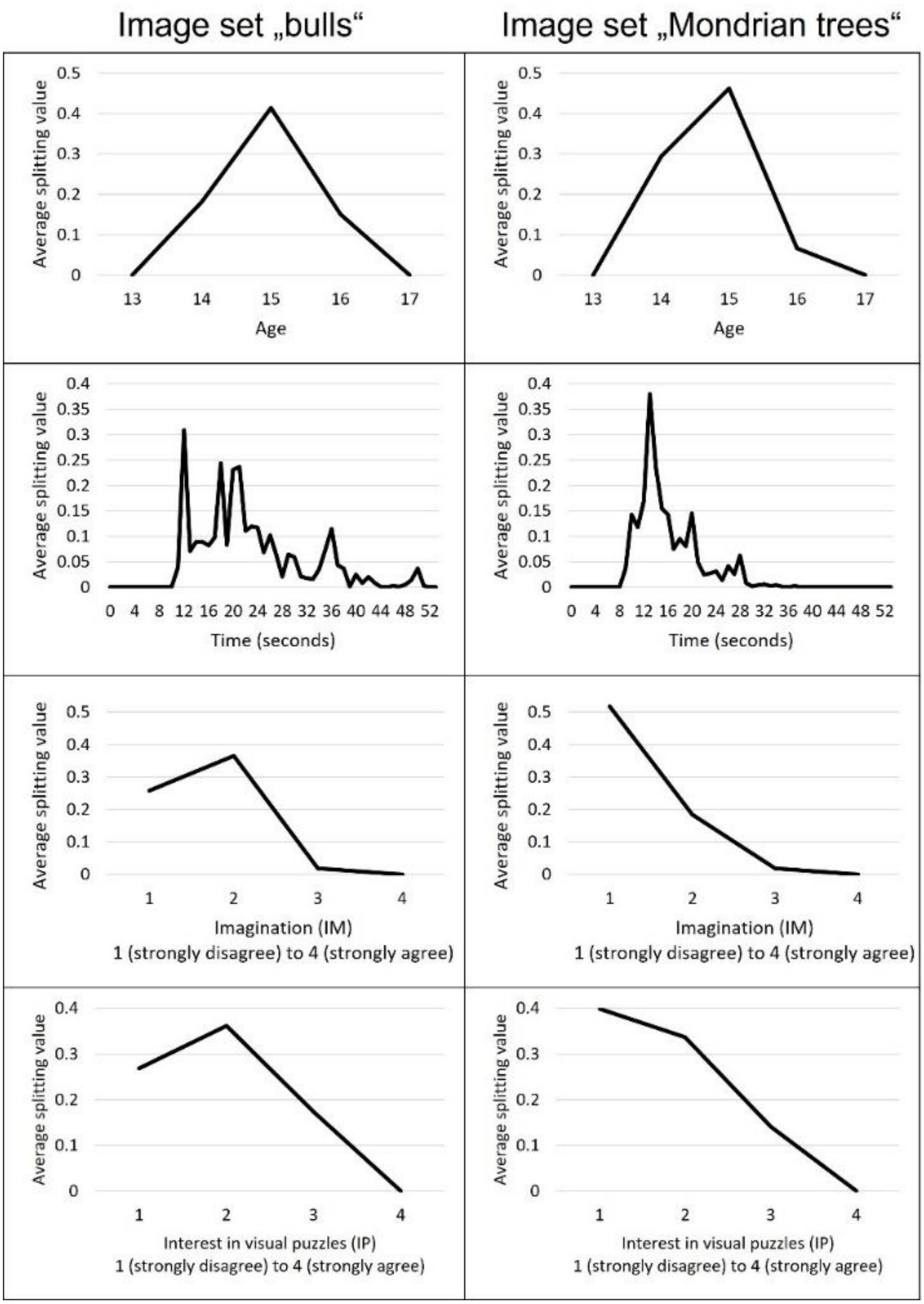
Figure 6 shows at which values continuous variables split the tree structure as a result of the bootstrapping procedure exemplified for the “bull images” and “Mondrian trees” image set in sample I. For the variable age most splits occurred for students above or below the age of 15 years. The time spent on the task varied for the bull images with a tendency to split at 5 seconds or between the 10-15 seconds. Whereas for the “Mondrian trees” splitting peaked around the 7-second mark and then continuously dropped until reaching zero at around 22 seconds.

Expert-Novice Comparison in Sample II

Worth parameters for the expert and novice comparison are listed in Table 5. Generally, experts showed a steeper, linear decline in worth parameters than novices. Subjects could not be grouped based on the number of fixations and the fixation duration on AOIs. Further, age was the only significant splitting variable on the “bull images” set.

Figure 6

Splitting Value for Continuous Variables in Sample I



Note. Average splitting values for the variables age, time, imagination (IM), and interest in visual puzzles (IP) on “bull images” (left) and “Mondrian trees” (right) as a result of the bootstrapping procedure in sample I.

Table 5

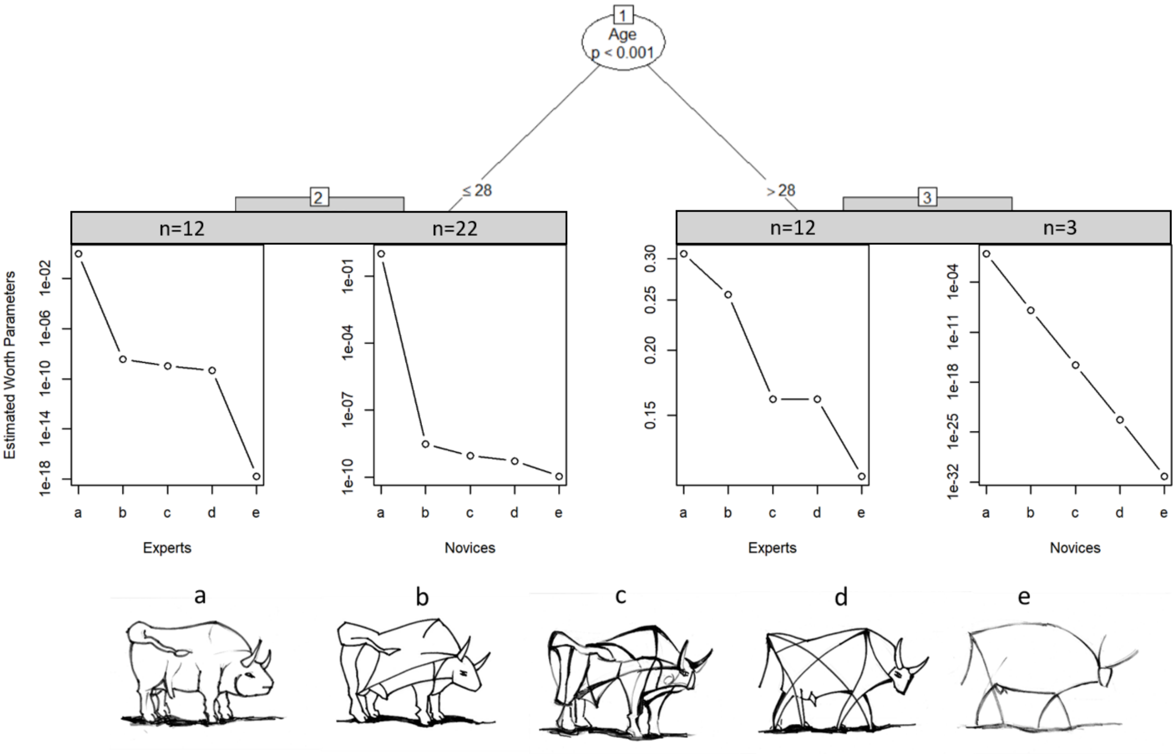
Worth Parameters in each Terminal Node from Sample II

Sample II - VL experts and novices (n=49)							
Image set	Terminal node	Worth parameters (π) for each image					Splitting Covariates
		a	b	c	d	e	
Geometric figures	n=24 Experts	0.999	1.65E-09	4.65E-18	1.31E-26	2.17E-35	-
	n=25 Novices	0.608	0.248	0.089	0.043	0.012	
Dogs	n=24 Experts	0.325	0.255	0.191	0.135	0.095	-
	n=25 Novices	0.478	0.197	0.168	0.100	0.057	
Bull images	n=12 Experts	0.999	3.57E-09	1.05E-09	4.55E-10	1.62E-18	Age <= 28
	n=22 Novices	0.999	3.01E-09	9.28E-10	5.25E-10	1.1E-10	
	n=12 Experts	0.307	0.256	0.161	0.161	0.114	Age > 28
	n=3 Novices	0.999	1.22E-08	2.54E-16	5.2E-24	6.45E-32	
Mondrian Trees	n=24 Experts	0.748	0.141	0.085	0.021	0.006	-
	n=25 Novices	0.999	2.21E-08	1.10E-08	2.58E-09	9.61E-10	

We take a closer look at how this item was perceived by the experts and novices. MOB LLBT results in in Figure 7 indicate that experts above age 28 judge bull image “c” and “d” to be very close in level of abstraction. In contrast, novices above the age of 28 estimate all bulls to have the same distance of abstraction to each other, however this may be due to the small sample size of only 3 novices in node 3. On the other hand, younger experts show a clear distinction between the most realistic and most abstract bull image, but differentiate only marginally between the three bull images in the middle. Novices below the age of 29 only differentiate strongly between the most realistic bull image to the rest. Generally, older participants differentiate better between the images.

Figure 7

Partitioned Paired Comparison Tree with Estimated Worth Parameters for the Ranking Task "Bull Images" in Sample II

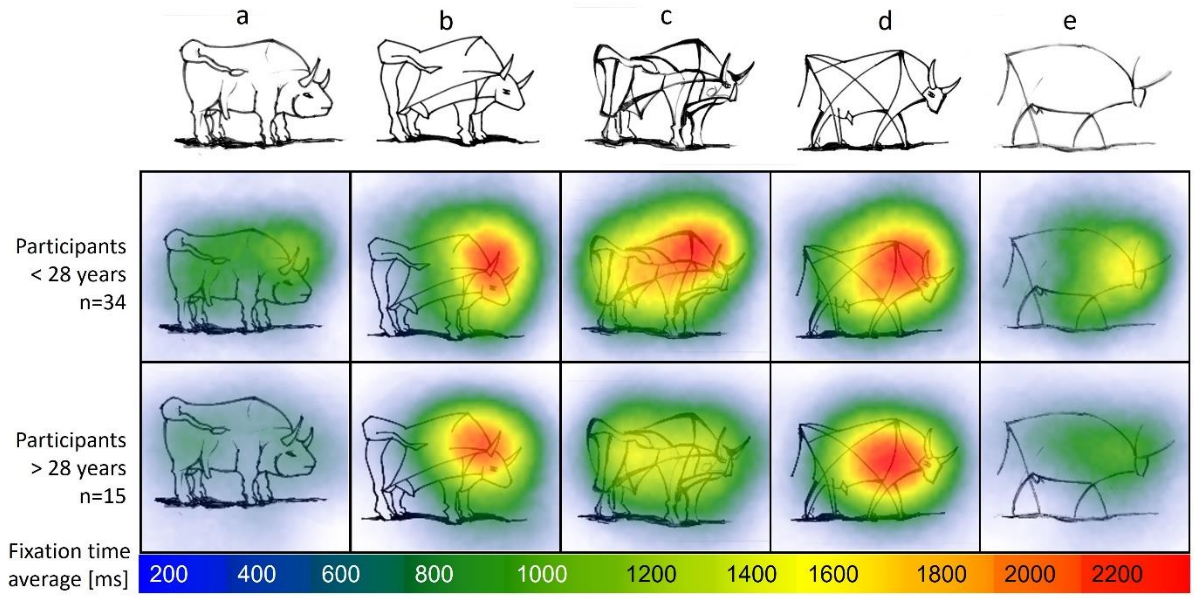


Next, we focus on the distribution of attention for the preference ranking through a fixation heatmap. The mean fixation time spent on the "bull image" set in sample II was $M_{Experts}$ = 18.37 sec (*SD* = 10.17), $M_{Novices}$ = 18.06 sec (*SD* = 8.38). Experts' and novices' fixation times did not significantly differ between each bull (F(4)=0.288, n.s.). A comparison of the distribution of fixations on each separate bull image during task completion revealed longer fixation times on bull images b, c and d compared to the most realistic (a) and most abstract (e) bull, F(4)=28.124, *p*<.001.

Figure 8 shows a heatmap of mean fixation durations on each bull AOI from start until end of trial, supplementing the model described in Fig. 7. The most abstract (right) and most realistic (left) bull image attract less attention compared to bulls of similar abstraction level. Fixation times of experts and novices was mainly spent on the bulls associated with a medium level of abstraction (b, c and d). There is a negative correlation between age and fixation time; *r*(47)= -.36, *p*= .011, i.e., older participants, spend less time on images compared to younger participants. Participants below 28 years spend additional fixation time on the most abstract bull image e compared to older groups.

Figure 8

Heatmap with Average Fixation Time on Image Set "Bull Images" by Age Groups



4 Discussion

This study explored how lay students, lay adults, and visual art experts ranked more or less abstract images by applying a LLBT model to identify potential heterogeneity in visual judgments. Overall, time to complete the ranking task in combination with self-reported skills have significant influence on model parameters. In general, the longer students took to rank the images, the closer each image was ranked to the previous one, i.e., the difference in the ranked preferences between the images decreases. Students who spent more time on the task had difficulties ranking the images the intended way. Additionally, visual skills affected the ease to differentiate between images. Interestingly, the students' art grade did not affect the ability to rank the presented images with respect to visual abstraction. There was also no apparent classroom group effect.

The slim packaging of the "salt packages" seems to determine the perceived difference in cost. In contrast to other images, the knowledge of goods and prices is very different to the evaluation of image abstraction and is well reflected by the preference scale: the divergence between small and round vs. slim and tall salt packaging can be clearly seen in the steep drop of estimated worth parameters after image "b". It could be hypothesized that male and female students might have different access to merchandise, which could explain the slight difference in cost perception by gender.

Furthermore, ranking abstract images such as the “bull images” revealed difficulties of students differentiating image pairs of similar abstraction level. A majority of students ranked bull image d as more realistic even though it contains less features than c. Apparently line thickness influences the perception of abstraction level for the majority of students. Also the bull’s eye is drawn slightly more realistically in bull d in comparison to bull c, which may have influenced the ranking. Are these differences in perceived judgment of images outside the intended ranking an indication for less skilled student groups? This cannot be derived solely from the ranked preferences. Comparing this result to the sample II, revealed how VL experts above the age of 28 judged both bull images c and d to be nearly identical in abstraction level. Exploring the fixation distribution of VL experts’ and novices’ eye movements, exemplified by heatmaps, showed how images of similar abstraction level (with similar worth parameters) evoke longer fixation durations.

Students with high self-reported interest in visual puzzle solving were able to distinguish abstract images more clearly. The self-reported ability to remember small details in pictures (“working memory”) also contributed to students’ ability to rank the level of abstraction of the images, indicated by greater systematic difference (i.e., exhibiting a steeper slope across the five images) in worth parameters between each image pair. Stability checks suggest that MOB LLBT models can sufficiently detect heterogeneity of visual judgments in a large sample of students. The time students took to rank the images was a significant splitting covariate for almost all image sets. The interest in visual puzzles was the most relevant self-reported ability for ranking abstract images. Furthermore, age, for example, was a less prevalent splitting variable for the “dogs” image set but not for the “bull images” and “Mondrian trees”. This might be caused by the difference between abstraction due to signal character (dogs as information) versus an aesthetic expression (trees and bulls as illustrations of experiences).

As seen in the results of the expert and novice comparison in sample II, VL experts were able to determine nuanced abstraction levels between images, as reflected in the similar worth parameters between image pairs. Smaller differences between certain image pairs do not necessarily reflect poorly on the ability to differentiate abstract images, but may indicate subtle image variations perceived by experts. Thus, especially when dealing with images of artwork, an interpretation by art experts and teachers is advisable.

In general, image sets used in the present study seemed to be of less difficulty for students. Judging images of more varying complexity (see García et al. (1994) for an early attempt to measure icon complexity) and difficulty could, thus, be a next step in the construction of future test batteries on VL. In contrast to measurements of visual abstraction with visual analog scales (e.g., the AAA instrument by Chatterjee et al. (2010)), ranking tasks lets participants compare multiple images at once. BT trees then can be used in various educational settings, e.g., art assignments where exact iconicity between two images is unknown. This modelling approach allows one to quantify the distance between images on a standardized latent scale. Here, BT models do not rely on the assumption of equidistant response categories. The latent metric scale is derived from ordinal (ranking) data to capture the perceived between-group differences of judgment. The perceived distance between each image (e.g., level of abstraction) can be used to identify closely related and, therefore, hard-to-differentiate objects. Such objects could subsequently be discussed and analyzed in art class.

As an empirically derived observation our results suggest the following: Time spent on task and ability to discriminate between images of varying levels of abstraction seem to go hand in hand. Abilities related to visual arts (imagination and interest in visual puzzles) seem to support this discriminative ability demonstrated by our participants.

Limitations

A few limitations of the present study should be mentioned. Firstly, as an exploratory study by design, generalizability of empirical results is limited. Only a reduced number of item sets were presented. Even though the intended ranking for abstract images was moderately low (between 29-42%), the images might have been too easy to solve, as no major outliers in worth parameters were observed across the student sample. Different sets of stimuli, e.g., computer generated art that controls for salience (Furnham & Rao, 2002; Shakeri et al., 2017) with a focus on a single dimensions of visual abstraction, such as composition or color (Markovic, 2010) could lead to higher variability in perceived judgment. In comparison to other image ranking tasks (e.g., Strobl et al., 2011) a ground truth exists (objectively correct ordering of items). However, a ranking assignment with heterogeneous preference patterns might indicate ambiguities with selected items. For educational assignments a clear preference ranking, with uniformly distributed worth parameters might be more desirable.

In sample II only age was found as a significant splitting variable, which might be due to low statistical power. Age of participants might also be confounded with expertise as older persons tend to have more expertise. Finally, the number of datapoints increase dramatically with the number of items for MOB LLBT models. With 5!=120 possible PC patterns and n=987 participants, the resulting input dataset consists of 118,440 observations, owing to the separate design matrices for each subject. Researchers might consider limiting the number of items during study design to reduce the design complexity.

5 Acknowledgments

We would like to thank all students, VL experts, and novices who participated in this study.

6 Declaration of Interest

All authors declare no financial interest or benefit that has arisen from the direct applications of this research. The results of this study were used by author MT to fulfil some of the requirements for the doctoral degree program at the University of Regensburg.

7 Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author.

8 Author Contributions

MT, UF and KR designed the study. MT and EW selected and prepared the stimuli. MT conducted the field work and eye-tracking experiments. MT performed the statistical analysis with input on data interpretation by WW, UF and MWG. MT prepared the manuscript. All authors reviewed the article and approved the submitted version.

9 Funding Details

This work was supported by the German Federal Ministry of Education and Research [BMBF] under grant number 01JK1606A.

CHAPTER 3. CONCLUSION

Summary

Goal of the presented studies was to explore VL sub-competencies and their effects on cognitive strategies during visual tasks. VL research from the perspective of cognitive psychology is interested in the perceptual abilities of VL and how specific differences in expertise can be empirically assessed. VL as operationalized by ENViL (Wagner & Schönau, 2016), comprises sixteen sub-competencies, three of which *interpreting*, *analyzing* and *judging* images were investigated further: three eye-tracking studies with VL experts and novices were conducted to test the validity of the assessment tool. Exploratory research was required to find empirical evidence of phenomenological assumptions with respect to VL. The measurement of eye movements recorded while participants solved visual tasks provided an indirect insight into experts' and laypeople's cognitive processes.

Previous research on domain specific visual expertise has found that experts can recover information from long-term memory more quickly, allocate attention more effectively, and encode features in greater chunks, as seen by their eye movement patterns (Brams et al., 2019). Our results show how VL experts take more and different areas of images into account than novices do. Detailed investigation into the sequential scanpaths (fixation sequences) between task-relevant image areas during visual search suggests expertise-specific solution strategies. Empirical assessment further revealed different student profiles on how to manage and solve visual tasks. Significant differences in students' abilities can be seen in all three studies, i.e., the comprehension of logical process visualizations, visual search and image rankings.

A deeper understanding of the cognitive processes involved in VL assessment tasks were assessed by eye tracking and by means of latent statistical modelling. Particularly the use of eye movements as spatio-temporal data in combination with latent modelling, as shown with HMMs, are a remarkably effective way to visualize and clarify the latent cognitive processes involved in visual problem solving. The idea to model eye movements as random variables underlying a stochastic process opens up a great range of sophisticated applications, particularly ones derived from the field of statistical machine learning (Boccignone, 2019). The use of such models are not only a suitable technique for eye movement event detection (e.g., Zhu et al., 2020) but also it is promising for the classification of top-down expertise-driven perceptual strategies. The following section reflects on the results in more detail.

The comprehension of visual logical models. A grouping algorithm such as LCA appears to be suitable for differentiating students based on their ability to decipher PMs. A total of six LCs differentiated between varying levels of comprehension of specific model parts. Beyond very good performers ("logic champions") and quite poor performers ("under performers") there exist other groups of students at intermediate levels, which can be related to qualitatively differing misunderstandings of the employed PMs. For example, some student groups only comprehend simultaneous activities in process models (LC2), others lack the comprehension of parallel paths (LC5), yet others cannot compare more than 2 relevant facts in one model (LC3). We assume the presence of a high motivation level across both samples as most students did not report any issues regarding understanding and following the task instructions, which supports the interpretation of the obtained LCs as *cognitive styles*. Knowing the solution profile of students allows teachers to provide meaningful feedback and develop appropriate techniques for the improvement of model comprehension.

Even though the PMs are displayed visually, the VL ability to *interpret* does not appear to aid in the understanding of the PM's logical structure. In terms of discriminatory validity, this result is useful in relation to other VL assessment items. A greater focus may be placed on tasks related to artistic expertise or aesthetic judgment. Nevertheless, it would be interesting to test if more artistic model notations, such as the inclusion of colors and special fonts, would make it easier for VL experts to grasp the logical character of PMs. It appears plausible however that the ability to solve PMs does not contribute to the distinctiveness of VL, highlighting an important distinction between visual logical models and other types of visual information. It seems that the search for subjective factors impacting PM comprehension should not address primarily VL but should rather be concerned by other cognitive capacities, e.g., intelligence or expertise in computer science and programming.

Recent studies have further explored the use of eye tracking to measure PM comprehension (Duarte et al., 2021). A next step may include the use of eye movement modeling examples (EMME) (Jarodzka et al., 2017), i.e., to show the recorded eye movements of experts to guide novices' attention throughout a visual model (Winter et al., 2021).

Visual search on artworks analyzed with HMMs. HMM results show how VL experts and novices differ in their attention allocation during visual search. Experts appear to pay more attention to smaller image locations, showing higher precision for some targets indicated by HMM state and AOI overlap. Fixation density maps generate appealing visualizations for the analysis of the distribution of attention. As a result, they are a valuable tool for exploratory investigations as well as a successful method for data-driven AOI identification. One significant limitation of traditional heatmaps is that they are restricted to the spatial dispersion of fixations at the expense of information about the order and temporal sequence of scanpaths (Bojko, 2009). The density maps based on HMM are suitable for research interested in the temporal aspects of eye movements. Additionally, the transition probability between data-driven AOIs may be considered to determine in what order latent image characteristics are inspected.

HMMs also allow for an aggregated comparison of multiple fixation sequences. Generally, comparing several scanpaths between individuals is a difficult task (Fahimi & Bruce, 2021). The threshold values, for example, to set a starting or landing point for fixations on key image areas is hard to determine a priori. When HMM are based on (expertise) groups the scanpaths can be derived through the transition probabilities between hidden states over time. For HMMs each fixation point can be categorized to its most likely hidden state. These hidden states can be understood as data-driven AOIs. Rather than establishing AOIs using arbitrary thresholds, we may allow subject fixations in whitespace (WS) that are outside the pre-defined AOIs to be included in any data-driven AOI (hidden state) depending on the estimated likelihood of belonging to a given state.

The use of HMM presented here is not exhaustive for the analysis of eye movements to capture expertise differences. Various possibilities are considerable and may be expanded upon: one possibility is the use of additional variables such as saccadic amplitudes or individual fixation durations, that might be useful to measure attention distribution. Currently a single fixation is not differentiated further regardless of whether it was an intermediate fixation to a target region or a visual scrutiny lasting several seconds. Weighting individual fixations (i.e., the emission probabilities of HMMs) may further refine the model.

Another option would be the use of *semi*-Markov Models (Yu, 2010). This approach would relax the *Markov property* of HMM by incorporating a memory for visual search. For example, a more realistic search behavior should consider an already found target region and “switch” to a different transition probability matrix after a successful partial search. A similar approach has been attempted by Chuk et al. (2019) to model cognitive state changes during facial perception. Furthermore, HMM can be used to investigate more complex search processes that can go beyond traditional visual search tasks (Haji-Abolhassani & Clark, 2013) and may be useful for more natural scenes and real-world searches, for example, for the analysis of visual quality inspection operations (Ulutas et al., 2019).

The judgment of visual abstraction. The results on the judgment of visual abstraction show that the amount of time spent on the ranking task, the students' age, and their self-reported interest in visual puzzles had a substantial impact on perceived judgment. Experts' and novices' fixation times revealed that both groups paid more attention to more difficult image pairs. The MOB LLBT model allows us to quantify the distance between perceived image characteristics on a standardized latent scale. This scale is constructed from ordinal (ranking) data in order to represent perceived variations in judgments across groups. The stability tests suggest that MOB LLBT models can detect visual judgment variability in a large sample of students. The apparent distance between each image (the degree of visual abstraction) can be utilized to detect items that are closely related and hence difficult to distinguish. Following that, such artworks could be discussed and evaluated further in art class. The use of BT models to diagnose latent discriminatory ability (instead of preferences) is a possibility that has so far received little consideration from empirical educational assessment and thus, if developed further, could lead to forthcoming applications.

The time spent on the task and the ability to distinguish between images of varying levels of abstraction appear to be linked. Proficiencies in the visual arts, such as the ability to *imagine* and an *interest in visual puzzles,* appear to support this discriminative ability. Additionally, minor differences in worth parameters between image pairs do not necessarily signify an inability to distinguish abstract images, but they may show subtle visual alterations recognized by experts.

Finally, art judgment may go beyond a single characteristic of an image and is often concerned by the overall quality or aesthetic value of artwork (Augustin & Leder, 2006; Chong, 2013; Winston & Cupchik, 1992). Nevertheless, the presented approach can also be used for these higher-level judgments of artwork when constructed as a PC task.

Limitations

Some limiting factors and restrains to generalizability of the results merit further discussion.

Eye-tracking equipment. As the eye-tracking experiments were primarily interested in the analysis of fixations on AOIs, we chose a head-mounted eye-tracker with a sampling frequency of 60Hz for flexibility of use and external validity (Appendix A): undergraduates underwent the same data gathering procedure as participants in the expert and novice groups with the exemption of wearing eye-tracking glasses. Other equipment could be considered for screen based visual tasks (e.g., a remote eye-tracker with chin rest). However, because subjects needed to look down on the tablet screen and touch it with their finger to input their answers we decided against a remote eye-tracking setup with unrestrained participants because of off sight data loss when participants move their heads (Niehorster et al., 2018).

When fixating AOIs from a dynamic visual angle, the likelihood of AOI-fixation detection errors increases (Orquin & Holmqvist, 2018). As a result, our AOIs were drawn more conservatively (larger) to compensate for uncertainty in eye tracking. In future investigations, using remote devices with constant lighting conditions and a steady head position, minimizing the pupil foreshortening effect (Hayes & Petrov, 2016), could eliminate this imprecision while simultaneously allowing pupillometric analysis. Considering that sampling frequency of 60Hz is low for modern eye-trackers it is still sufficient for fixation-based eye movement analysis and has been used in other studies implementing HMM during visual search (Haji-Abolhassani & Clark, 2013). The use of a higher sampling frequency may still improve data quality (Andersson et al., 2010).

Eye-tracking technology is in rapid development (Holmqvist et al., 2011; Holmqvist & Andersson, 2017; Orquin & Holmqvist, 2019) with improvements in precision and accuracy of data collection as well as ease of use. Future endeavors may consider eye movement analysis of complex joint-interactions such as in classroom lessons (Jarodzka et al., 2021).

Visual tasks and stimuli. To generalize the results of the presented visual tasks to the overall effect of VL would be an oversimplification. Only a small number of items were presented in each tasks untypical for perceptual learning tasks with a few hundred trials (Fine & Jacobs, 2002). In general, the image sets employed appeared to be easy to solve for undergraduate students (e.g., no major outliers in worth parameters were observed across the student sample for the ranking task). As a result, using more items of higher difficulty could be a further step in the development of future VL test batteries to avoid ceiling effects.

Furthermore, the images were not evaluated for low-level saliency (Foulsham, 2019). Different levels of saliency may have an impact on the visual search and ranking items (Loftus & Mackworth, 1978). Consequently, future research could systematically adjust the number of visual prominent elements, e.g., the use of a saliency map based on low-level perceptual features (Le Meur & Baccino, 2013) to distinguish between bottom-up cognitive states from top-down strategies used by VL experts.

Finally, the selection of items analyzed were mainly concerned with *receptive* features of VL (*interpreting, analyzing, judging*). *Producing* art is, however, just as important for visual art expertise (Chamberlain, 2018) and therefore be indicative for VL expertise. Such items, outside the scope of this dissertation, are included as assessment tasks and are considered in future assessments (Frick, Rakoczy, Tallon, Weiß, & Wagner, 2020).

Sample of VL experts and novices. The number of participants was above average for eye-tracking studies (Brams et al., 2019; Gegenfurtner et al., 2011) but the sample size might still be too low to find every difference between the VL experts and non-expert group. There was also an age difference between the groups and expertise might be confounded with age as older people tend to gain more expertise during their lifetime. However, the engagement with and the production of visual art is not necessarily common for older people; e.g., only 5% of adults above 52 years of age visit art galleries or museums on a monthly basis and over 50% visit museums never or less than once a year (Fancourt & Steptoe, 2018).

As a validated assessment instrument for VL is yet to be constructed, the group of VL experts were instead comprised from a variety of visual art domains (from art teacher to self-employed artists). The use of recently validated art interest and knowledge questionnaires, e.g., Specker et al. (2020), might be useful for future studies on VL to authenticate expertise status and make expert groups more comparable and the findings thereof more replicable. Even though art expertise can only be seen as a subcategory it could be used as an approximation of VL.

Outlook and Concluding Remarks

The presented studies make an empirical psychological contribution to the analysis of the *perceiving* sub-competencies of VL and offered multiple ways to evaluate perceptual abilities through expert-novice comparisons with probabilistic models and eye tracking.

The results of the visual tasks examined in this thesis are beneficial for both VL research and visual art expertise research. The CEFR-VL comprises low-level perceptual abilities as well as higher top-down cognitive abilities declared as sub-competencies of VL. It is therefore important to distinguish between these sub-competencies of VL. Results show how low-level perceptual abilities were important for the visual search and the ranking task of visual abstraction.

Empirical findings from brain and psychological research suggest enhanced cognitive abilities for visual experts, such as memory recognition (Evans et al., 2011) and an increase in attention through aesthetic appreciation (Sarasso et al., 2020). Seeley and Kozbelt (2008) argue that artists develop enhanced encoding of targets in their visual field via declarative knowledge of art compositions and their practice of artistic productive techniques. These attentional strategies may explain the perceptual advantages (Kozbelt & Seeley, 2007) of VL experts.

Research from empirical aesthetics further indicate that low-level perceptual abilities can distinguish art experts from novices but these differences may not transfer to other visual domains (Angelone et al., 2016; Evans et al., 2011; Jarodzka et al., 2010; Pang et al., 2013). Differentiations are mostly seen in the production of artwork, e.g., drawing ability (Calabrese & Marucci, 2006; Drake et al., 2021). The type of cognitive processing involved in the presented visual tasks are a combination of automatic, but also top-down processes as suggested by empirical models of art perception (Pelowski et al., 2017). Bottom-up processes can further be influenced by domain-specific art expertise and declarative knowledge (Pelowski et al., 2017; Vogt & Magnussen, 2007). The results demonstrate how experts and novices differ in the way they deconstruct and analyze visual imagery, i.e., the *process* of task completion is more indicative of VL expertise than the achieved outcome, e.g., the total number of correct answers.

However, most empirically observed differences between VL experts and novices are more subtle than that indicated by VL models. Experts and novices only marginally differed in their total number of correct targets discovered during visual search, and experts' judgment of visual abstraction is nuanced and refined for specific item pairs. Classical aggregate statistics may fall short when tasked to analyze spatio-temporal eye movement data. Subtle differences in viewing behavior can be more clearly defined by utilizing latent statistical models for classification of response patterns and eye movements. The presented research shows how new methodological approaches and multidisciplinary studies can advance the field of empirical VL research and indicate which visual tasks are more likely to differentiate between VL experts and non-experts.

Only one third of current VL research is conducted with quantitative methods, most of which only use survey data, and another third is conducted without any reported empirical methodology (Matusiak, 2020). The interdisciplinary field of VL-research cannot be thoroughly explored without a combination of psychometrics, art education research, and new methodological approaches such as eye tracking. Because the literature is dominated by studies in sports and medicine, further empirical research in the domain of arts is required, as eye movements may vary across visual domains (Brams et al., 2019).

As more empirical research is conducted on specific VL sub-competencies a stronger conclusion can be drawn on the significance of VL as an overarching latent construct. Today's research on VL has demonstrated that it is not a "failed metaphor" (Cassidy & Knowlton, 1983) as research strives to specify VL as measurable competency. The CEFR-VL encourages psychological empirical research to support this competency-oriented approach:

> The competency model [CEFR-VL] is based on both research in educational psychology on the development of knowledge and abilities, and empirical evidence in the field of *Visual Literacy* education, enabling an accessible, research-based model to be developed. (Wagner & Schönau, 2016, p. 100)

The CEFR-VL is in revision (Schönau et al., 2021; Schönau & Kárpáti, 2019) with adaptation to the number of sub-competencies and discussions on basic versus domain specific dimensions of VL. Leben (2019) points out how basic perceptual visual competencies (e.g., identifying the degree of image iconicity) could be separated by higher-order visual competencies that are also affected by cultural upbringing (e.g., when differentiating contemporary from traditional art). The results of the presented studies support the idea of differentiating VL sub-competencies from perceptual learning abilities and abilities found in visual art expertise. The community of VL research has grown into a diverse field with different scientific disciplines and methodological approaches (Brumberger, 2019; Levie, 1978; Matusiak, 2020; Michelson, 2017). Recent approaches in empirical VL and visual art expertise research are still in early development. As a result, more research is expected to expand the field, find empirical evidence of proposed (sub-)competencies and lay groundwork for a validated VL assessment tool. It is up to the research community to decide how empirical evidence of VL is taken into consideration for future model iterations to subsequently foster VL in art education.

Bibliography

Aguilar-Savén, R. S. (2004). Business process modelling: Review and framework. *International Journal of Production Economics*, *90*(2), 129–149. https://doi.org/10.1016/s0925-5273(03)00102-6

Allweyer, T. (2016). *BPMN 2.0: introduction to the standard for business process modeling*. BoD–Books on Demand.

Andersson, R., Larsson, L., Holmqvist, K., Stridh, M., & Nyström, M. (2017). One algorithm to rule them all? An evaluation and discussion of ten eye movement event-detection algorithms. *Behavior Research Methods*, *49*(2), 616–637. https://doi.org/10.3758/s13428-016-0738-9

Andersson, R., Nyström, M., & Holmqvist, K. (2010). Sampling frequency and eye-tracking measures: how speed affects durations, latencies, and more. *Journal of Eye Movement Research*, *3*(3), 1–12. https://doi.org/10.16910/JEMR.3.3.6

Andrà, C., Ferdinando, A., Francesca, F., & Kenneth, H. (2009). How students read mathematical representations: An eye tracking study. In M. Tzekaki, M. Kaldrimidou, & C. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education.* PME.

Andrews, K., Zimoch, M., Reichert, M., Tallon, M., Frick, U., & Pryss, R. (2018). A Smart Mobile Assessment Tool for Collecting Data in Large-Scale Educational Studies. *Procedia Computer Science*, *134*, 67–74. https://doi.org/10.1016/j.procs.2018.07.145

Angelone, B. L., Hass, R. W., & Cohen, M. (2016). Skill transfer in visual arts expertise. *Psychology of Aesthetics, Creativity, and the Arts*, *10*(2), 147–156. https://doi.org/10.1037/aca0000053

Antes, J. R., & Kristjanson, A. F. (1991). Discriminating artists from nonartists by their eye-fixation patterns. *Perceptual and Motor Skills*, *73*(3 Pt 1), 893–894. https://doi.org/10.2466/pms.1991.73.3.893

Arendasy, M. E., & Sommer, M. (2013). Reducing response elimination strategies enhances the construct validity of figural matrices. *Intelligence*, *41*(4), 234–243.

Asparouhov, T., & Muthén, B. (2012). Using Mplus TECH11 and TECH14 to test the number of latent classes. *Mplus Web Notes*, *14*, 22.

Augustin, M. D., & Leder, H. (2006). Art expertise: A study of concepts and conceptual spaces. *Psychology Science*, *48*, 135–157.

Augustin, M. D., Leder, H., Hutzler, F., & Carbon, C.-C. (2008). Style follows content: On the microgenesis of art perception. *Acta Psychologica, 128*(1), 127–138. https://doi.org/10.1016/j.actpsy.2007.11.006

Avgerinou, M. D. (2003). A mad-tea party no-more: Revisiting the visual literacy definition problem. In R. E. Griffin, V. S. Williams, & L. Jung (Eds.), *Turning trees* (pp. 29–41). Loretto, PA: IVLA.

Avgerinou, M. D., & Pettersson, R. (2011). Toward a Cohesive Theory of Visual Literacy. *Journal of Visual Literacy*, *30*(2), 1–19. https://doi.org/10.1080/23796529.2011.11674687

Bačić, D., & Fadlalla, A. (2016). Business information visualization intellectual contributions: An integrative framework of visualization capabilities and dimensions of visual intelligence. *Decision Support Systems, 89*, 77–86. https://doi.org/10.1016/j.dss.2016.06.011

Bahle, B., Thayer, D. D., Mordkoff, J. T., & Hollingworth, A. (2019). The architecture of working memory: Features from multiple remembered objects produce parallel, coactive guidance of attention in visual search. *Journal of Experimental Psychology. General.* Advance online publication. https://doi.org/10.1037/xge0000694

Bauer, D. (2014). *Meaning Making with Art:: Expert and Lay Perspectives in Understanding Artworks and Exhibition Concepts* [PhD Thesis, Universität Tübingen]. repositorium.ixtheo.de. https://repositorium.ixtheo.de/xmlui/handle/10900/56823

Becker, J., Rosemann, M., & Uthmann, C. von. (2000). Guidelines of Business Process Modeling. In *Business Process Management* (pp. 30–49). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45594-9_3

Bednarik, R., & Tukiainen, M. (2006). An eye-tracking methodology for characterizing program comprehension processes. In K.-J. Räihä & A. T. Duchowski (Eds.) (pp. 125–132). ACM.

Bentwich, M. E., & Gilbey, P. (2017). More than visual literacy: Art and the enhancement of tolerance for ambiguity and empathy. *BMC Medical Education*, *17*(1), 200. https://doi.org/10.1186/s12909-017-1028-7

Berger, I., Shamir, A., Mahler, M., Carter, E., & Hodgins, J. (2013). Style and abstraction in portrait sketching. *ACM Transactions on Graphics, 32*(4), 1–12. https://doi.org/10.1145/2461912.2461964

Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning: Componential and eye movement analysis of geometric analogy performance. *Intelligence*, *8*(3), 205–238.

Bharathi, S., Chervenak, A., Deelman, E., Mehta, G., Su, M.-H., & Vahi, K. (2008). Characterization of scientific workflows. In *2008 Third Workshop on Workflows in Support of Large-Scale Science*. IEEE. https://doi.org/10.1109/works.2008.4723958

Blunsom, P. (2004). *Hidden markov models*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.1016&rep=rep1&type=pdf

Boccignone, G. (2019). Advanced Statistical Methods for Eye Movement Analysis and Modelling: A Gentle Introduction. In C. Klein & U. Ettinger (Eds.), *Eye movement research: An introduction to its scienctific foundations and applications* (Vol. 6, pp. 309–405). Springer. https://doi.org/10.1007/978-3-030-20085-5_9

Boisvert, J. F., & Bruce, N. D. (2016). Predicting task from eye movements: On the importance of spatial distribution, dynamics, and image features. *Neurocomputing*, *207*, 653–668. https://doi.org/10.1016/j.neucom.2016.05.047

Bojko, A. (2009). Informative or Misleading? Heatmaps Deconstructed. In J. A. Jacko (Ed.), *Lecture Notes in Computer Science: Vol. 5610. New trends* (Vol. 5610, pp. 30–39). Springer. https://doi.org/10.1007/978-3-642-02574-7_4

Borji, A., & Itti, L. (2013). State-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(1), 185–207. https://doi.org/10.1109/TPAMI.2012.89

Borji, A., & Itti, L. (2014). Defending Yarbus: Eye movements reveal observers' task. *Journal of Vision*, *14*(3), 29. https://doi.org/10.1167/14.3.29

Bourgeois, A., Chelazzi, L., & Vuilleumier, P. (2016). How motivation and reward learning modulate selective attention. *Progress in Brain Research*, *229*, 325–342. https://doi.org/10.1016/bs.pbr.2016.06.004

Boy, J., Rensink, R. A., Bertini, E., & Fekete, J. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1963–1972. https://doi.org/10.1109/tvcg.2014.2346984

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324–345. https://doi.org/10.2307/2334029

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

Brill, J. M., Kim, D., & Branch, R. (2007). Visual Literacy Defined – The Results of a Delphi Study: Can IVLA (Operationally) Define Visual Literacy? *Journal of Visual Literacy*, *27*(1), 47–60. https://doi.org/10.1080/23796529.2007.11674645

Brown, I. M., & Lockyer, L. (2007). *Don't Mention the 'T' Word: Value and Validity in Testing Visual Literacy* (Selected Readings of the International Visual Literacy Association). Loretto, PA. The International Visual Literacy Association.

Brumberger, E. (2011). Visual Literacy and the Digital Native: An Examination of the Millennial Learner. *Journal of Visual Literacy*, *30*(1), 19–47. https://doi.org/10.1080/23796529.2011.11674683

Brumberger, E. (2019). Past, present, future: mapping the research in visual literacy. *Journal of Visual Literacy*, *38*(3), 165–180. https://doi.org/10.1080/1051144X.2019.1575043

Brumberger, E. (2021). The potential of eye tracking for visual literacy research. *Journal of Visual Literacy*, *40*(1), 34–50. https://doi.org/10.1080/1051144X.2021.1902040

Bucher, H.-J., & Schumacher, P. (2006). The relevance of attention for selecting news content. An eye-tracking study on attention patterns in the reception of print and online media. *Communications*, *31*(3), 347–368.

Calabrese, L., & Marucci, F. S. (2006). The influence of expertise level on the visuo-spatial ability: Differences between experts and novices in imagery and drawing abilities. *Cognitive Processing*, *7*(S1), 118–120. https://doi.org/10.1007/s10339-006-0094-2

Cassidy, M. F., & Knowlton, J. Q. (1983). Visual Literacy: A Failed Metaphor? *Educational Communication and Technology*, *31*(2), 66–90. http://www.jstor.org/stable/30218097

Cattelan, M. (2012). Models for Paired Comparison Data: A Review with Emphasis on Dependent Data. *Statistical Science*, *27*(3), 412–433. https://doi.org/10.1214/12-sts396

Cattelan, M., Varin, C., & Firth, D. (2013). Dynamic Bradley-Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(1), 135–150. https://doi.org/10.1111/j.1467-9876.2012.01046.x

Chakraborty, H., & Sen, P. K. (2016). Resampling method to estimate intra-cluster correlation for clustered binary data. *Communications in Statistics - Theory and Methods*, *45*(8), 2368–2377. https://doi.org/10.1080/03610926.2013.870202

Chamberlain, R. (2018). The Development of Expertise in the Visual Arts. In G. Campitelli, D. Z. Hambrick, & B. N. Macnamara (Eds.), *Frontiers of Cognitive Psychology. The science of expertise: Behavioral, neural, and genetic approaches to complex skill* (pp. 129–150). Routledge. https://doi.org/10.4324/9781315113371-8

Chamberlain, R., Drake, J. E., Kozbelt, A., Hickman, R., Siev, J., & Wagemans, J. (2019). Artists as experts in visual cognition: An update. *Psychology of Aesthetics, Creativity, and the Arts*, *13*(1), 58–73. https://doi.org/10.1037/aca0000156

Chamorro-Premuzic, T., & Furnham, A. (2004). Art Judgment: A Measure Related to Both Personality and Intelligence? *Imagination, Cognition and Personality*, *24*(1), 3–24. https://doi.org/10.2190/U4LW-TH9X-80M3-NJ54

Chan, C. Y. H., Chan, A. B., Lee, T. M. C., & Hsiao, J. H. (2018). Eye-movement patterns in face recognition are associated with cognitive decline in older adults. *Psychonomic Bulletin & Review*, *25*(6), 2200–2207. https://doi.org/10.3758/s13423-017-1419-0

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*(8), 1146–1152. https://doi.org/10.3758/bf03206384

Chatterjee, A., Widick, P., Sternschein, R., Smith, W. B., & Bromberger, B. (2010). The Assessment of Art Attributes. *Empirical Studies of the Arts*, *28*(2), 207–222. https://doi.org/10.2190/EM.28.2.f

Chen, Y.-C., & Yang, F.-Y. (2014). Probing the relationship between process of spatial problems solving and science learning: An eye tracking approach. *International Journal of Science and Mathematics Education*, *12*(3), 579–603.

Child, I. L. (1965). Personality correlates of esthetic judgment in college students. *Journal of Personality*, *33*(3), 476–511. https://doi.org/10.1111/j.1467-6494.1965.tb01399.x

Choi, B. C. K., & Pak, A. W. P. (2005). A catalog of biases in questionnaires. *Preventing Chronic Disease*, *2*(1), A13.

Chong, P. (2013). Legitimate judgment in art, the scientific world reversed? Maintaining critical distance in evaluation. *Social Studies of Science*, *43*(2), 265–281. https://doi.org/10.1177/0306312712475256

Chuk, T., Chan, A. B., & Hsiao, J. H. (2014). Understanding eye movements in face recognition using hidden Markov models. *Journal of Vision*, *14*(11). https://doi.org/10.1167/14.11.8

Chuk, T., Chan, A. B., Shimojo, S., & Hsiao, J. H. (2019). Eye movement analysis with switching hidden Markov models. *Behavior Research Methods.* Advance online publication. https://doi.org/10.3758/s13428-019-01298-y

Coutrot, A., Hsiao, J. H., & Chan, A. B. (2018). Scanpath modeling and classification with hidden Markov models. *Behavior Research Methods*, *50*(1), 362–379. https://doi.org/10.3758/s13428-017-0876-8

Cox, J. A., & Aimola Davies, A. M. (2020). Keeping an eye on visual search patterns in visuospatial neglect: A systematic review. *Neuropsychologia*, 107547. https://doi.org/10.1016/j.neuropsychologia.2020.107547

Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Attention, Perception, & Psychophysics*, *83*(5), 2241–2255. https://doi.org/10.3758/s13414-021-02268-3

David, H. A. (1988). *The method of paired comparisons* (2. ed., rev). *Griffin's statistical monographs and courses: Vol. 41*. Oxford Univ. Press; Griffin.

Dayton, C. M., & Macready, G. B. (2006). Latent Class Analysis in Psychometrics. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (Vol. 26, pp. 421–446). Elsevier. https://doi.org/10.1016/S0169-7161(06)26013-9

Derek Panchuk, Samuel Vine, & Joan N. Vickers. (2015). Eye Tracking Methods in Sport Expertise. In *Routledge international handbooks. Routledge handbook of sport expertise* (pp. 176–187). Routledge. https://doi.org/10.4324/9781315776675-16

Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *47*(4), 511–525. https://doi.org/10.1111/1467-9876.00125

Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (2002). Modelling dependencies in paired comparison data. *Computational Statistics & Data Analysis*, *40*(1), 39–57. https://doi.org/10.1016/S0167-9473(01)00106-2

Dittrich, R., Katzenbeisser, W., & Reisinger, H. (2000). The analysis of rank ordered preference data based on Bradley-Terry Type Models. *OR Spectrum*, *22*(1), 117–134. https://doi.org/10.1007/s002910050008

Drake, J. E., Riccio, A., Chamberlain, R., & Kozbelt, A. (2021). Artists have superior local and global processing abilities but show a preference for initially drawing globally. *Psychology of Aesthetics, Creativity, and the Arts.* Advance online publication. https://doi.org/10.1037/aca0000424

Drew, T., Evans, K., Võ, M. L.-H., Jacobson, F. L., & Wolfe, J. M. (2013). Informatics in radiology: What can you see in a single glance and how might this guide visual search in medical images? *Radiographics : A Review Publication of the Radiological Society of North America, Inc*, *33*(1), 263–274. https://doi.org/10.1148/rg.331125023

Duarte, R. B., Da Silveira, D. S., Albuquerque Brito, V. de, & Lopes, C. S. (2021). A systematic literature review on the usage of eye-tracking in understanding process models. *Business Process Management Journal*, *27*(1), 346–367. https://doi.org/10.1108/BPMJ-05-2020-0207

Dumas, M., La Rosa, M., Mendling, J., Mäesalu, R., Reijers, H. A., & Semenenko, N. (2012). Understanding business process models: the costs and benefits of structuredness. In J. Ralyté, X. Franch, S. Brinkkemper, & S. Wrycza (Eds.) (pp. 31–46). Springer.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2013). *Fundamentals of business process management*. Springer.

Engbert, R. (2006). Microsaccades: a microcosm for research on oculomotor control, attention, and visual perception. In *Visual Perception - Fundamentals of Vision: Low and Mid-Level Processes in Perception* (pp. 177–192). Elsevier. https://doi.org/10.1016/s0079-6123(06)54009-9

Engbert, R., Rothkegel, L. O. M., Backhaus, D., & Trukenbrod, H. A. (2016). *Evaluation of velocity-based saccade detection in the SMI-ETG 2W system.* Technical Report.

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

Ericsson, K. A., & Towne, T. J. (2010). Expertise. *WIREs Cognitive Science*, *1*(3), 404–416. https://doi.org/10.1002/wcs.47

Evans, K. K., Cohen, M. A., Tambouret, R., Horowitz, T., Kreindel, E., & Wolfe, J. M. (2011). Does visual expertise improve visual recognition memory? *Attention, Perception & Psychophysics*, *73*(1), 30–35. https://doi.org/10.3758/s13414-010-0022-5

Fahimi, R., & Bruce, N. D. B. (2021). On metrics for measuring scanpath similarity. *Behavior Research Methods*, *53*(2), 609–628. https://doi.org/10.3758/s13428-020-01441-0

Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, *33*(3), 397–412. https://doi.org/10.1016/0042-6989(93)90094-d

Fahle, M., & Poggio, T. (2002). *Perceptual learning*. MIT Press.

Fan, J. E., Hawkins, R. D., Wu, M., & Goodman, N. D. (2020). Pragmatic Inference and Visual Abstraction Enable Contextual Flexibility During Visual Communication. *Computational Brain & Behavior, 3*(1), 86–101. https://doi.org/10.1007/s42113-019-00058-7

Fancourt, D., & Steptoe, A. (2018). Cultural engagement predicts changes in cognitive function in older adults over a 10 year period: Findings from the English Longitudinal Study of Ageing. *Scientific Reports*, *8*(1), 10226. https://doi.org/10.1038/s41598-018-28591-8

Ferguson, S. L., G. Moore, E. W., & Hull, D. M. (2020). Finding latent groups in observed data: A primer on latent profile analysis in Mplus for applied researchers. *International Journal of Behavioral Development, 44*(5), 458–468. https://doi.org/10.1177/0165025419881721

Ferretti, G., Mazzotti, S., & Brizzolara, D. (2008). Visual scanning and reading ability in normal and dyslexic children. *Behavioural Neurology*, *19*(1-2), 87–92. https://doi.org/10.1155/2008/564561

Figl, K. (2017). Comprehension of procedural visual business process models. *Business & Information Systems Engineering, 59*(1), 41–67.

Figl, K., Mendling, J., & Strembeck, M. (2013). The Influence of Notational Deficiencies on Process Model Comprehension. *Journal of the Association for Information Systems*, *14*(6), 312–338. https://doi.org/10.17705/1jais.00335

Finan, N. (2002). Visual literacy in images used for medical education and health promotion. *The Journal of Audiovisual Media in Medicine, 25*(1), 16–23. https://doi.org/10.1080/0140511022011837X

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. *Journal of Vision*, *2*(2), 190–203. https://doi.org/10.1167/2.2.5

Formann, A. K., Waldherr, K., & Piswanger, K. (2011). *Wiener Matrizen-Test 2: WMT-2. Ein Rasch-skalierter sprachfreier Kurztest zu Erfassung der Intelligenz.* Hogrefe.

Foulsham, T. (2019). Scenes, Saliency Maps and Scanpaths. In C. Klein & U. Ettinger (Eds.), *Eye movement research: An introduction to its scienctific foundations and applications* (Vol. 25, pp. 197–238). Springer. https://doi.org/10.1007/978-3-030-20085-5_6

Fox, S. E., & Faulkner-Jones, B. E. (2017). Eye-Tracking in the Study of Visual Expertise: Methodology and Approaches in Medicine. *Frontline Learning Research*, *5*(3), 29–40. https://eric.ed.gov/?id=ej1148377

Franceschini, S., Gori, S., Ruffino, M., Pedrolli, K., & Facoetti, A. (2012). A causal link between visual spatial attention and reading acquisition. *Current Biology : CB*, *22*(9), 814–819. https://doi.org/10.1016/j.cub.2012.03.013

Francis, B., Dittrich, R., & Hatzinger, R. (2010). Modeling heterogeneity in ranked responses by nonparametric maximum likelihood: How do Europeans get their scientific knowledge? *Annals of Applied Statistics*, *4*(4), 2181–2202. https://doi.org/10.1214/10-AOAS366

Francuz, P., Zaniewski, I., Augustynowicz, P., Kopiś, N., & Jankowski, T. (2018). Eye Movement Correlates of Expertise in Visual Arts. *Frontiers in Human Neuroscience*, *12*, 87. https://doi.org/10.3389/fnhum.2018.00087

Frank, S. M., Greenlee, M. W., & Tse, P. U. (2018). Long Time No See: Enduring Behavioral and Neuronal Changes in Perceptual Learning of Motion Trajectories 3 Years After Training. *Cerebral Cortex*, *28*(4), 1260–1271. https://doi.org/10.1093/cercor/bhx039

Frick, U., Rakoczy, K., Tallon, M., & Weiß, S. (2020). *Bildkompetenz in der Kulturellen Bildung (BKKB) - Was ist und wie fördert man Bildkompetenz? : Schlussbericht : Förderkennzeichen "BKKB Assessment": 01JK1606A, Förderzeitraum: 11/2016-0172019 : Förderkennzeichen "BKKB Unterricht": 01JK1606B, Förderzeitraum: 11/2016-01/2020*. Technische Informationsbibliothek (TIB). https://doi.org/10.2314/KXP:1759261807

Frick, U., Rakoczy, K., Tallon, M., Weiß, S., & Wagner, E. (2020). Ich sehe was, was Du nicht siehst! Erste Bausteine zur Messung von Bildkompetenz bei Schüler*innen der 9. und 10. Jahrgangsstufe [I can see what you cannot see! First building blocks for measuring visual literacy in 9th and 10th grade]. In S. Timm, J. Costa, C. Kühn, & A. Scheunpflug (Eds.), *Kulturelle Bildung: Theoretische Perspektiven, methodologische Herausforderungen und empirische Befunde.* Waxmann.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. https://doi.org/10.1016/j.socec.2010.10.008

Furnham, A., & Rao, S. (2002). Personality and the Aesthetics of Composition: A study of Mondrian and Hirst. *North American Journal of Psychology*(4), 233–242.

García, M., Badre, A. N., & Stasko, J. T. (1994). Development and validation of icons varying in their abstractness. *Interacting with Computers*, *6*(2), 191–211. https://doi.org/10.1016/0953-5438(94)90024-8

Gartus, A., Völker, M., & Leder, H. (2020). What Experts Appreciate in Patterns: Art Expertise Modulates Preference for Asymmetric and Face-Like Patterns. *Symmetry*, *12*(5), 707. https://doi.org/10.3390/sym12050707

Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, *113*, 212–225. https://doi.org/10.1016/j.compedu.2017.06.001

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

Gegenfurtner, A., & van Merriënboer, J. J. G. (2017). Methodologies for studying visual expertise. *Frontline Learning Research*, *5*(3), 1–13. https://doi.org/10.14786/flr.v5i3.316

Gold, J. I., & Watanabe, T. (2010). Perceptual learning. *Current Biology : CB*, *20*(2), R46-8. https://doi.org/10.1016/j.cub.2009.10.066

Gortais, B. (2003). Abstraction and art. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *358*(1435), 1241–1249. https://doi.org/10.1098/rstb.2003.1309

Graf, L. K. M., & Landwehr, J. R. (2017). Aesthetic Pleasure versus Aesthetic Interest: The Two Routes to Aesthetic Liking. *Frontiers in Psychology*, *8*(373), 214. https://doi.org/10.3389/fpsyg.2017.00015

Greenlee, M. W. (2014). The Neuronal Base of Perceptual Learning and Skill Acquisition. In S. Billett, C. Harteis, & H. Gruber (Eds.), *Springer International Handbooks of Education. International Handbook of Research in Professional and Practice-based Learning* (pp. 305–336). Springer Netherlands. https://doi.org/10.1007/978-94-017-8902-8_12

Greenlee, M. W., Frank, S. M., Reavis, E. A., & Tse, P. U. (2014). Individual differences in perceptual learning: structural and functional brain correlates. *Perception*, *43*(Suppl1), 120. https://pred.uni-regensburg.de/id/eprint/9852

Groenendijk, T., Haanstra, F., & Kárpáti, A. (2018). A new tool for developmental assessment based on the Common European Framework of Reference for Visual Literacy – an international usability study. *International Journal of Education Through Art*, *14*(3), 353–378. https://doi.org/10.1386/eta.14.3.353_1

Groenendijk, T., Kárpáti, A., & Haanstra, F. (2020). Self-assessment in Art Education through a Visual Rubric. *International Journal of Art & Design Education*, *39*(1), 153–175. https://doi.org/10.1111/jade.12233

Gruhn, V., & Laue, R. (2006). Adopting the Cognitive Complexity Measure for Business Process Models. In *2006 5th IEEE International Conference on Cognitive Informatics,* Beijing, China.

Guinibert, M. (2020). Learn from your environment: A visual literacy learning model. *Australasian Journal of Educational Technology*, *36*(4), 173–188. https://doi.org/10.14742/ajet.5200

Güney, Z. (2019). Visual Literacy and Visualization in Instructional Design and Technology for Learning Environments. *European Journal of Contemporary Education, 8*(1), 103–117. https://eric.ed.gov/?id=ej1212285

Hagenaars, J. A., & McCutcheon, A. L. (2002). *Applied Latent Class Analysis*. Cambridge University Press.

Haider, H., & Frensch, P. A. (1996). The Role of Information Reduction in Skill Acquisition. *Cognitive Psychology*, *30*(3), 304–337. https://doi.org/10.1006/cogp.1996.0009

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

Haji-Abolhassani, A., & Clark, J. J. (2013). A computational model for task inference in visual search. *Journal of Vision*, *13*(11). https://doi.org/10.1167/13.3.29

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse Yarbus process: Predicting observers' task from eye movement patterns. *Vision Research*, *103*, 127–142. https://doi.org/10.1016/j.visres.2014.08.014

Harris, J. (2014). *Sensation and Perception*. Sage.

Hatzinger, R., & Dittrich, R. (2012). prefmod : An R Package for Modeling Preferences Based on Paired Comparisons, Rankings, or Ratings. *Journal of Statistical Software, 48*(10). https://doi.org/10.18637/jss.v048.i10

Hayes, T. R., & Petrov, A. A. (2016). Mapping and correcting the influence of gaze position on pupil size measurements. *Behavior Research Methods*, *48*(2), 510–527. https://doi.org/10.3758/s13428-015-0588-x

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision, 11*(10), 10.

Hayn-Leichsenring, G. U., Kenett, Y. N., Schulz, K., & Chatterjee, A. (2020). Abstract art paintings, global image properties, and verbal descriptions: An empirical and computational investigation. *Acta Psychologica*, *202*, 102936. https://doi.org/10.1016/j.actpsy.2019.102936

Hekkert, P., & van Wieringen, P. C. (1996). The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica*, *94*(2), 117–131. https://doi.org/10.1016/0001-6918(95)00055-0

Hogrebe, F., Gehrke, N., & Nüttgens, M. (2011). Eye Tracking Experiments in Business Process Modeling: Agenda Setting and Proof of Concept. In EMISA (pp). In M. Nüttgens, O. Thomas, & B. Weber (Eds.) (pp. 183–188). Gesellschaft für Informatik e.V.

Hollingworth, A., & Bahle, B. (2020). Eye Tracking in Visual Search Experiments. In S. Pollmann (Ed.), *Neuromethods: Vol. 151. Spatial Learning and Attention Guidance* (1st ed., Vol. 151, pp. 23–35). Springer US; Imprint: Humana. https://doi.org/10.1007/7657_2019_30

Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). CreateSpace.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press USA - OSO. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=5825719

Hothorn, T., & Zeileis, A. (2015). partykit: A Modular Toolkit for Recursive Partytioning in R. *Journal of Machine Learning Research*, *16*, 3905–3909. https://jmlr.org/papers/v16/hothorn15a.html

Houpt, J. W., Frame, M. E., & Blaha, L. M. (2018). Unsupervised parsing of gaze data with a beta-process vector auto-regressive hidden Markov model. *Behavior Research Methods*, *50*(5), 2074–2096. https://doi.org/10.3758/s13428-017-0974-7

Hsiao, J. H., An, J., Zheng, Y., & Chan, A. B. (2021). Do portrait artists have enhanced face processing abilities? Evidence from hidden Markov modeling of eye movements. *Cognition*, *211*, 104616. https://doi.org/10.1016/j.cognition.2021.104616

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, 105–134. https://doi.org/10.4324/9780203488430

Jacobsen, T. (2004). Individual and group modelling of aesthetic judgment strategies. *British Journal of Psychology (London, England : 1953)*, *95*(Pt 1), 41–56. https://doi.org/10.1348/000712604322779451

Jarodzka, H., Gruber, H., & Holmqvist, K. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, *10(1)*(3), 1–18.

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*(2), 146–154. https://doi.org/10.1016/j.learninstruc.2009.02.019

Jarodzka, H., Skuballa, I., & Gruber, H. (2021). Eye-Tracking in Educational Practice: Investigating Visual Perception Underlying Teaching and Learning in the Classroom. *Educational Psychology Review*, *33*(1), 1–10. https://doi.org/10.1007/s10648-020-09565-7

Kappas, A., & Olk, B. (2008). The concept of visual competence as seen from the perspective of the psychological and brain sciences. *Visual Studies*, *23*(2), 162–173. https://doi.org/10.1080/14725860802276313

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250–252. https://doi.org/10.1038/365250a0

Kaya, M. (2020). The Impact of Visual Literacy Awareness Education on Verbal and Writing Skills of Middle School Students. *International Journal of Education and Literacy Studies*, *8*(2), 71. https://doi.org/10.7575/aiac.ijels.v.8n.2p.71

Kazmierczak, E. T. (2001). Iconicity, diagrammatics, and aesthetic preferences: A semiotic perspective on visual literacy and information design. *Visual Sociology*, *16*(1), 89–99. https://doi.org/10.1080/14725860108583828

Kędra, J. (2018). What does it mean to be visually literate? Examination of visual literacy definitions in a context of higher education. *Journal of Visual Literacy*, *37*(2), 67–84. https://doi.org/10.1080/1051144X.2018.1492234

Kellman, P. J., & Garrigan, P. (2009). Perceptual learning and human expertise. *Physics of Life Reviews*, *6*(2), 53–84. https://doi.org/10.1016/j.plrev.2008.12.001

Koide, N., Kubo, T., Nishida, S., Shibata, T., & Ikeda, K. (2015). Art expertise reduces influence of visual salience on fixation in viewing abstract-paintings. *PloS One*, *10*(2), e0117696. https://doi.org/10.1371/journal.pone.0117696

Kolodziej, M., Majkowski, A., Francuz, P., Rak, R., & Augustynowicz, P. (2018). Identifying experts in the field of visual arts using oculomotor signals, *3*(11), 1–10. https://doi.org/10.16910/jemr.11.3.3

Komogortsev, O. V., Gobert, D. V., Jayarathna, S., Koh, D.-H., & Gowda, S. (2010). Standardization of automated analyses of oculomotor fixation and saccadic behaviors. *IEEE Transactions on Bio-Medical Engineering*, *57*(11), 2635–2645. https://doi.org/10.1109/tbme.2010.2057429

Koochaki, F., & Najafizadeh, L. (Eds.). (2018). *Predicting Intention Through Eye Gaze Patterns*. https://doi.org/10.1109/BIOCAS.2018.8584665

Kozbelt, A., & Seeley, W. P. (2007). Integrating art historical, psychological, and neuroscientific explanations of artists' advantages in drawing and perception. *Psychology of Aesthetics, Creativity, and the Arts*, *1*(2), 80–90. https://doi.org/10.1037/1931-3896.1.2.80

Kundel, H. L., Nodine, C. F., Conant, E. F., & Weinstein, S. P. (2007). Holistic component of image perception in mammogram interpretation: Gaze-tracking study. *Radiology*, *242*(2), 396–402. https://doi.org/10.1148/radiol.2422051997

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90–115.

Latto, R., Brain, D., & Kelly, B. (2000). An oblique effect in aesthetics: Homage to Mondrian (1872-1944). *Perception*, *29*(8), 981–987. https://doi.org/10.1068/p2352

Laurence, P. G., Mecca, T. P., Serpa, A., Martin, R., & Macedo, E. C. (2018). Eye Movements and Cognitive Strategy in a Fluid Intelligence Test: Item Type Analysis. *Frontiers in Psychology*, *9*, 380.

Le Meur, O., & Baccino, T. (2013). Methods for comparing scanpaths and saliency maps: Strengths and weaknesses. *Behavior Research Methods*, *45*(1), 251–266. https://doi.org/10.3758/s13428-012-0226-9

Leben, L. (2019). Visuelle Komptenzen messen: Mit Testungen Lernerfolge thematisieren [Measuring Visual Competency - Addressing learning success with tests]. *BDK-Mitteilungen*(3), 22–26. https://www.bdk-online.info/blog/2019/08/09/bdk-mitteilungen-3-2019/

Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*(95), 489–508.

Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition & Emotion*, *28*(6), 1137–1147. https://doi.org/10.1080/02699931.2013.870132

Leder, H., Gerger, G., Dressler, S. G., & Schabmann, A. (2012). How art is appreciated. *Psychology of Aesthetics, Creativity, and the Arts*, *6*(1), 2–10. https://doi.org/10.1037/a0026396

Leder, H., & Nadal, M. (2014). Ten years of a model of aesthetic appreciation and aesthetic judgments : The aesthetic episode - Developments and challenges in empirical aesthetics. *British Journal of Psychology (London, England : 1953)*, *105*(4), 443–464. https://doi.org/10.1111/bjop.12084

Levie, W. H. (1978). A Prospectus for Instructional Research on Visual Literacy. *Educational Communication and Technology*, *26*(1), 25–36. http://www.jstor.org/stable/30217957

Li, R., Pelz, J., Shi, P., Alm, C. O., & Haake, A. R. (2012). Learning eye movement patterns for characterization of perceptual expertise. In C. H. Morimoto, H. Istance, J. B. Mulligan, P. Qvarfordt, & S. N. Spencer (Eds.), *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '12* (p. 393). ACM Press. https://doi.org/10.1145/2168556.2168645

Li, W., & Gilbert, C. D. (2008). 2.20 - Perceptual Learning. In R. H. Masland, T. D. Albright, P. Dallos, D. Oertel, S. Firestein, G. K. Beauchamp, M. Catherine Bushnell, A. I. Basbaum, J. H. Kaas, & E. P. Gardner (Eds.), *The Senses: A Comprehensive Reference* (pp. 303–328). Academic Press. https://doi.org/10.1016/B978-012370880-9.00333-9

Li, W., & Gilbert, C. D. (2009). Perceptual Learning: Neural Mechanisms. In L. R. Squire (Ed.), *Encyclopedia of neuroscience* (pp. 535–541). Academic Elsevier. https://doi.org/10.1016/B978-008045046-9.00227-8

Loerts, T., & Belcher, C. (2019). Developing visual literacy competencies while learning course content through visual journaling: teacher candidate perspectives. *Journal of Visual Literacy*, *38*(1-2), 46–65. https://doi.org/10.1080/1051144X.2018.1564603

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology. Human Perception and Performance*, *4*(4), 565–572. https://doi.org/10.1037//0096-1523.4.4.565

Luce, R. D. (1959). *Individual Choice Behavior: A theoretical analysis, New York, NY: John Willey and Sons.*

MacTaggart, J. (2021, May 4). *Animals in Art - Pablo Picasso*. https://www.artyfactory.com/art_appreciation/animals_in_art/pablo_picasso.htm

Markovic, S. (2010). Perceptual, semantic and affective dimensions of the experience of representational and abstract paintings. *Journal of Vision*, *10*(7), 1230. https://doi.org/10.1167/10.7.1230

Martinez-Conde, S., Macknik, S. L., Troncoso, X. G., & Hubel, D. H. (2009). Microsaccades: A neurophysiological analysis. *Trends in Neurosciences*, *32*(9), 463–475. https://doi.org/10.1016/j.tins.2009.05.006

Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: Towards a unified theory of saccadic function. *Nature Reviews Neuroscience*, *14*(2), 83–96. https://doi.org/10.1038/nrn3405

Mason, L., Scheiter, K., & Tornatora, M. C. (2017). Using eye movements to model the sequence of text-picture processing for multimedia comprehension. *Journal of Computer Assisted Learning*, *33*(5), 443–460. https://doi.org/10.1111/jcal.12191

Matusiak, K. K. (2020). Studying visual literacy: Research methods and the use of visual evidence. *IFLA Journal*, *46*(2), 172–181. https://doi.org/10.1177/0340035219886611

Mayer, R. E. (1996). Learning strategies for making sense out of expository text: The SOI model for guiding three cognitive processes in knowledge construction. *Educational Psychology Review*, *8*(4), 357–371.

Mayer, R. E. (1999). Designing instruction for constructivist learning. *Instructional-Design Theories and Models: A New Paradigm of Instructional Theory*, *2*, 141–159.

McCormack, J., Cruz Gambardella, C., & Lomas, A. (2021). The Enigma of Complexity. In J. Romero, T. Martins, & N. Rodríguez-Fernández (Eds.), *Lecture Notes in Computer Science: Vol. 12693. Artificial Intelligence in Music, Sound, Art and Design* (Vol. 12693, pp. 203–217). Springer International Publishing. https://doi.org/10.1007/978-3-030-72914-1_14

McCutcheon, A. (Ed.). (1987). *Latent Class Analysis*. SAGE Publications, Inc. https://doi.org/10.4135/9781412984713

McManus, I. C., Cheema, B., & Stoker, J. (1993). The Aesthetics of Composition: A Study of Mondrian. *Empirical Studies of the Arts*, *11*(2), 83–94. https://doi.org/10.2190/HXR4-VU9A-P5D9-BPQQ

McMaster, S. (2015). *Visual Literacy and Art Education: Review of the Literature.* https://doi.org/10.13140/RG.2.1.4331.2482

Meißner, M., & Oll, J. (2017). The Promise of Eye-Tracking Methodology in Organizational Research: A Taxonomy, Review, and Future Avenues. *Organizational Research Methods*, *22*(2), 590–617. https://doi.org/10.1177/1094428117744882

Mendling, J., Strembeck, M., & Recker, J. (2012). Factors of process model comprehension—Findings from a series of experiments. *Decision Support Systems*, *53*(1), 195–206. https://doi.org/10.1016/j.dss.2011.12.013

Michelson, A. (2017). A short history of visual literacy: the first five decades. *Art Libraries Journal*, *42*(2), 95–98. https://doi.org/10.1017/alj.2017.10

Moody, D. L., Sindre, G., Brasethvik, T., & Sølvberg, A. (2002). Evaluating the Quality of Process Models: Empirical Testing of a Quality Framework. In (pp. 380–396). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-45816-6_36

Mor, B., Garhwal, S., & Kumar, A. (2021). A Systematic Review of Hidden Markov Models and Their Applications. *Archives of Computational Methods in Engineering*, *28*(3), 1429–1448. https://doi.org/10.1007/s11831-020-09422-4

Muhammad, U. R., Yang, Y., Song, Y.-Z., Xiang, T., & Hospedales, T. M. (2018). Learning Deep Sketch Abstraction. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* IEEE. https://doi.org/10.1109/cvpr.2018.00836

Muller-Gass, A., Duncan, M., & Campbell, K. (2017). Brain states predict individual differences in perceptual learning. *Personality and Individual Differences*, *118*, 29–38. https://doi.org/10.1016/j.paid.2017.03.066

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature*, *434*(7031), 387.

Nakamura, C., & Zeng-Treitler, Q. (2012). A Taxonomy of Representation Strategies in Iconic Communication. *International Journal of Human-Computer Studies*, *70*(8), 535–551. https://doi.org/10.1016/j.ijhcs.2012.02.009

Netzer, O., Ebbes, P., & Bijmolt, T. H. A. (2017). Hidden Markov Models in Marketing. In P. S. H. Leeflang, J. E. Wieringa, T. H. A. Bijmolt, & K. Pauwels (Eds.), *International Series in Quantitative Marketing. Advanced methods for modeling markets* (Vol. 10, pp. 405–449). Springer. https://doi.org/10.1007/978-3-319-53469-5_14

Nodine, C. F., Carmody, D. P., & Herman, E. (1979). Eye movements during visual search for artistically embedded targets. *Bulletin of the Psychonomic Society*, *13*(6), 371–374. https://doi.org/10.3758/BF03336898

Nodine, C. F., Locher, P. J., & Krupinski, E. A. (1993). The Role of Formal Art Training on Perception and Aesthetic Judgment of Art Compositions. *Leonardo*, *26*(3), 219–227.

Oberski, D. (2016). Mixture Models: Latent Profile and Latent Class Analysis. In *Modern Statistical Methods for HCI* (pp. 275–287). Springer, Cham. https://doi.org/10.1007/978-3-319-26633-6_12

Olivers, C. N. L., Huettig, F., Singh, J. P., & Mishra, R. K. (2013). The influence of literacy on visual search. *Visual Cognition*, *22*(1), 74–101. https://doi.org/10.1080/13506285.2013.875498

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

OMG, Object Management Group. (2011). *Business Process Model and Notation (BPMN) Version 2.0*. https://www.omg.org/spec/BPMN/2.0/

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods*, *50*(4), 1645–1656. https://doi.org/10.3758/s13428-017-0998-z

Orquin, J. L., & Holmqvist, K. (2019). A Primer on Eye-Tracking Methodology for Behavioral Science. In *A Handbook of Process Tracing Methods* (pp. 53–64). Routledge. https://doi.org/10.4324/9781315160559-5

Pang, C. Y., Nadal, M., Müller-Paul, J. S., Rosenberg, R., & Klein, C [C.] (2013). Electrophysiological correlates of looking at paintings and its association with art expertise. *Biological Psychology*, *93*(1), 246–254. https://doi.org/10.1016/j.biopsycho.2012.10.013

Pelowski, M., Markey, P. S., Forster, M., Gerger, G., & Leder, H. (2017). Move me, astonish me… delight my eyes and brain: The Vienna Integrated Model of top-down and bottom-up processes in Art Perception (VIMAP) and corresponding affective, evaluative, and neurophysiological correlates. *Physics of Life Reviews*, *21*, 80–125. https://doi.org/10.1016/j.plrev.2017.02.003

Petrusel, R., & Mendling, J. (2013). Eye-tracking the factors of process model comprehension tasks. In S. Camille, N. Moira C., & P. Óscar (Eds.), *Advanced Information Systems Engineering* (Vol. 7908, pp. 224–239). Springer Berlin Heidelberg.

Philipp, M., Rusch, T., Hornik, K., & Strobl, C. (2018). Measuring the Stability of Results From Supervised Statistical Learning. *Journal of Computational and Graphical Statistics*, *27*(4), 685–700. https://doi.org/10.1080/10618600.2018.1473779

Pihko, E., Virtanen, A., Saarinen, V.-M., Pannasch, S., Hirvenkari, L., Tossavainen, T., Haapala, A., & Hari, R. (2011). Experiencing art: The influence of expertise and painting abstraction level. *Frontiers in Human Neuroscience*, *5*, 94. https://doi.org/10.3389/fnhum.2011.00094

Piras, A., Lobietti, R., & Squatrito, S. (2014). Response time, visual search strategy, and anticipatory skills in volleyball players. *Journal of Ophthalmology*, *2014*, 189268. https://doi.org/10.1155/2014/189268

Punzalan, J. F. (2018). The impact of visual arts in students' academic performance, *6*, 121–130. https://www.ijern.com/journal/2018/july-2018/10.pdf

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Rakoczy, K., Frick, U., Weiß, S., Tallon, M., & Wagner, E. (2019). Was ist und wie fördert man Bildkompetenz? Ansätze und Methoden der Wahrnehmungspsychologie und der quantitativ-empirischen Bildungsforschung [What is and how do you promote Visual Literacy? Approaches and methods of perceptual psychology and quantitative-empirical educational research]. *BDK-Mitteilungen*(3), 18–21. https://www.bdk-online.info/blog/2019/08/09/bdk-mitteilungen-3-2019/

Recker, J., & Dreiling, A. (2007). *Does it matter which process modelling language we teach or use? An experimental study on understanding process modelling languages without formal education*. https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1045&context=acis2007

Recker, J., & Dreiling, A. (2011). The Effects of Content Presentation Format and User Characteristics on Novice Developers' Understanding of Process Models. *Communications of the Association for Information Systems*, *28*. https://doi.org/10.17705/1CAIS.02806

Reggio, G., Ricca, F., Scanniello, G., Di Cerbo, F., & Dodero, G. (2015). On the comprehension of workflows modeled with a precise style: results from a family of controlled experiments. *Software & Systems Modeling*, *14*(4), 1481–1504. https://doi.org/10.1007/s10270-013-0386-9

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, *12*(1), 48–55. https://doi.org/10.1111/1467-9280.00309

Reingold, E. M., & Sheridan, H. (2011). Eye movements and visual expertise in chess and medicine. *Oxford Handbook on Eye Movements*, *528*.

Roehm, T., Tiarks, R., Koschke, R., & Maalej, W. (2012). How do professional developers comprehend software? In *34th International Conference on Software Engineering (ICSE).* IEEE. https://doi.org/10.1109/icse.2012.6227188

Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*, *61*, 224–236. https://doi.org/10.1016/j.jbi.2016.04.007

Rost, J., & Langeheine, R. (1997). A guide through latent structure models for categorical data. *Applications of Latent Trait and Latent Class Models in the Social Sciences*, 13–37.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the symposium on Eye tracking research & applications - ETRA '00.* ACM Press. https://doi.org/10.1145/355017.355028

Sarasso, P., Ronga, I., Kobau, P., Bosso, T., Artusio, I., Ricci, R., & Neppi-Modona, M. (2020). Beauty in mind: Aesthetic appreciation correlates with perceptual facilitation and attentional amplification. *Neuropsychologia*, *136*, 107282. https://doi.org/10.1016/j.neuropsychologia.2019.107282

Schabmann, A., Gerger, G., Schmidt, B. M., Wögerer, E., Osipov, I., & Leder, H. (2016). Where Does It Come From? Developmental Aspects of Art Appreciation. *International Journal of Behavioral Development*, *40*(4), 313–323. https://doi.org/10.1177/0165025415573642

Schepman, A., & Rodway, P. (2021). Shared meaning in representational and abstract visual art: An empirical study. *Psychology of Aesthetics, Creativity, and the Arts*, *15*(3), 458–469. https://doi.org/10.1037/aca0000279

Schepman, A., Rodway, P., Pullen, S. J., & Kirkham, J. (2015). Shared liking and association valence for representational art but not abstract art. *Journal of Vision*, *15*(5), 11. https://doi.org/10.1167/15.5.11

Schoch, K., & Ostermann, T. (2020). Giving the Art Greater weight in Art Psychology: Rizba, a Questionnaire for Formal Picture Analysis. *Creativity. Theories – Research - Applications*, *7*(2), 373–410. https://doi.org/10.2478/ctra-2020-0019

Schönau, D., & Kárpáti, A. (2019). Renaming the framework: Common European Framework of Reference for Visual Competency. *International Journal of Education Through Art*, *15*(1), 95–100. https://doi.org/10.1386/eta.15.1.95_1

Schönau, D., Kárpáti, A., & Kirchner, C. (2021). *New Structural Model of Visual Competencies in Visual Literacy: The Revised Common European Framework of*

Reference for Visual Competency. European … ENVIL. http://envil.eu/wp-content/uploads/2021/06/04_Schonau_Karpati_Kirchner_Letsiou.pdf

Schultheiss, L. A., & Heiliger, E. M. (1963). Techniques of flow-charting. *0069-4789*. http://www.ideals.illinois.edu/handle/2142/743

Seeley, W., & Kozbelt, A. (2008). Art, Artists, and Perception: A Model for Premotor Contributions to Perceptual Analysis and Form Recognition. *Philosophical Psychology*, *21*(2), 149–171. https://doi.org/10.1080/09515080801976573

Seitz, A. R. (2017). Perceptual learning. *Current Biology : CB*, *27*(13), R631-R636. https://doi.org/10.1016/j.cub.2017.05.053

Shakeri, H., Nixon, M., & DiPaola, S. (2017). Saliency-Based Artistic Abstraction With Deep Learning and Regression Trees. *Journal of Imaging Science and Technology, 61*(6), 60402. https://doi.org/10.2352/J.ImagingSci.Technol.2017.61.6.060402

Sheridan, H., & Reingold, E. M. (2014). Expert vs. Novice differences in the detection of relevant information during a chess game: Evidence from eye movements. *Frontiers in Psychology*, *5*, 941. https://doi.org/10.3389/fpsyg.2014.00941

Sheridan, H., & Reingold, E. M. (2017). The Holistic Processing Account of Visual Expertise in Medical Image Perception: A Review. *Frontiers in Psychology*, *8,* 1620. https://doi.org/10.3389/fpsyg.2017.01620

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428. https://doi.org/10.1037//0033-2909.86.2.420

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Silvia, P. J. (2006). Artistic Training and Interest in Visual Art: Applying the Appraisal Model of Aesthetic Emotions. *Empirical Studies of the Arts*, *24*(2), 139–161. https://doi.org/10.2190/DX8K-6WEA-6WPA-FM84

Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, *9*(4), 237–251. https://doi.org/10.1016/j.cogsys.2008.01.002

Sinclair, C. D. (1982). GLIM for Preference. In D. Brillinger, S. Fienberg, J. Gani, J. Hartigan, K. Krickeberg, & R. Gilchrist (Eds.), *Lecture Notes in Statistics. GLIM 82: Proceedings of the International Conference on Generalised Linear Models* (Vol. 14, pp. 164–178). Springer New York. https://doi.org/10.1007/978-1-4612-5771-4_16

Solso, R. L. (2001). Brain Activities in a Skilled versus a Novice Artist: An fMRI Study. *Leonardo*, *34*(1), 31–34. https://doi.org/10.1162/002409401300052479

Spalter, A. M., & van Dam, A. (2008). Digital Visual Literacy. *Theory into Practice*, *47*(2), 93–101. https://doi.org/10.1080/00405840801992256

Specht, S. M. (2007). Successive Contrast Effects for Judgments of Abstraction in Artwork following Minimal Pre-Exposure. *Empirical Studies of the Arts*, *25*(1), 63–70. https://doi.org/10.2190/W717-88W2-2233-12H3

Specker, E., Forster, M., Brinkmann, H., Boddy, J., Pelowski, M., Rosenberg, R., & Leder, H. (2020). The Vienna Art Interest and Art Knowledge Questionnaire (VAIAK): A unified and validated measure of art interest and art knowledge. *Psychology of Aesthetics, Creativity, and the Arts*, *14*(2), 172–185. https://doi.org/10.1037/aca0000205

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, *59*(1), 1–34. https://doi.org/10.1348/000711005X48266

Stokes, S. (2002). *Visual literacy in teaching and learning: A literature perspective*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.136.7248&rep=rep1&type=pdf

Streiner, D. L., & Norman, G. R. (2008). Biases in responding. In D. L. Streiner & G. R. Norman (Eds.), *Health measurement scales: A practical guide to their development and use* (4th ed., pp. 103–134). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199231881.003.0006

Strobl, C., Wickelmaier, F., & Zeileis, A. (2011). Accounting for Individual Differences in Bradley-Terry Models by Means of Recursive Partitioning. *Journal of Educational and Behavioral Statistics*, *36*(2), 135–153. https://doi.org/10.3102/1076998609359791

Stuart, S., Hickey, A., Vitorio, R., Welman, K., Foo, S., Keen, D., & Godfrey, A. (2019). Eye-tracker algorithms to detect saccades during static and dynamic tasks: A structured review. *Physiological Measurement*, *40*(2), 02TR01. https://doi.org/10.1088/1361-6579/ab02ab

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 19–30). Cambridge University Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). Measuring Cognitive Load. In J. Sweller, P. Ayres, & S. Kalyuga (Eds.), *Cognitive Load Theory* (pp. 71–85). Springer New York. https://doi.org/10.1007/978-1-4419-8126-4_6

Sweller, J., & Sweller, S. (2006). Natural Information Processing Systems. *Evolutionary Psychology*, *4*(1), 147470490600400135. https://doi.org/10.1177/147470490600400135

Tallon, M., Greenlee, M. W., Wagner, E., Rakoczy, K., & Frick, U. (2021). How Do Art Skills Influence Visual Search? - Eye Movements Analyzed With Hidden Markov Models. *Frontiers in Psychology*, *12*, 594248. https://doi.org/10.3389/fpsyg.2021.594248

Tallon, M., Winter, M., Pryss, R., Rakoczy, K., Reichert, M., Greenlee, M. W., & Frick, U. (2019). Comprehension of business process models: Insight into cognitive strategies via eye tracking. *Expert Systems with Applications*, *136*, 145–158. https://doi.org/10.1016/j.eswa.2019.06.032

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, *34*(4), 273–286. https://doi.org/10.1037/h0070288

Tien, T., Pucher, P. H., Sodergren, M. H., Sriskandarajah, K., Yang, G.-Z., & Darzi, A. (2014). Eye tracking for skills assessment and training: A systematic review. *The Journal of Surgical Research*, *191*(1), 169–178. https://doi.org/10.1016/j.jss.2014.04.032

Tijus, C., Barcenilla, J., Lavalette, B. C., & Meunier, J.-G. (2007). The design, understanding and usage of pictograms. In D. Alamargot, P. Terrier, & J. M. Cellier (Eds.), *Written documents in the workplace* (pp. 17–31). Elsevier. https://doi.org/10.1163/9789004253254_003

Tillmann, A. (2012). *What We See and Why It Matters: How Competency in Visual Literacy can Enhance Student Learning* (No. 9). Illinois Wesleyan University. https://digitalcommons.iwu.edu/education_honproj/9

Tutz, G., & Schauberger, G. (2015). Extended ordered paired comparison models with application to football data from German Bundesliga. *AStA Advances in Statistical Analysis*, *99*(2), 209–227. https://doi.org/10.1007/s10182-014-0237-1

Tyler, C. W., & Likova, L. T. (2012). The role of the visual arts in the enhancing the learning process. *Frontiers in Human Neuroscience*, *6*, 8. https://doi.org/10.3389/fnhum.2012.00008

Ulutas, B. H., Özkan, N. F., & Michalski, R. (2019). Application of hidden Markov models to eye tracking data analysis of visual quality inspection operations. *Central European Journal of Operations Research*, *207*(11), 653. https://doi.org/10.1007/s10100-019-00628-x

Ungan, M. (2006). Towards a better understanding of process documentation. *The TQM Magazine*, *18*(4), 400–409. https://doi.org/10.1108/09544780610671066

Vaeyens, R., Lenoir, M., Williams, A. M., Mazyn, L., & Philippaerts, R. M. (2007). The effects of task constraints on visual search behavior and decision-making skill in youth soccer players. *Journal of Sport & Exercise Psychology*, *29*(2), 147–169. https://doi.org/10.1123/jsep.29.2.147

Vakil, E., & Lifshitz-Zehavi, H. (2012). Solving the Raven Progressive Matrices by adults with intellectual disability with/without Down syndrome: different cognitive patterns as indicated by eye-movements. *Research in Developmental Disabilities*, *33*(2), 645–654.

Vaknin-Nusbaum, V., & Nevo, E. (2021). What can children learn from exposure to visual-graphical representations? *Thinking Skills and Creativity*, *40*, 100830. https://doi.org/10.1016/j.tsc.2021.100830

van der Gijp, A., Ravesloot, C. J., Jarodzka, H., van der Schaaf, M. F., van der Schaaf, I. C., van Schaik, J. P. J., & Cate, T. J. ten (2017). How visual search relates to visual diagnostic performance: A narrative systematic review of eye-tracking research in radiology. *Advances in Health Sciences Education : Theory and Practice*, *22*(3), 765–787. https://doi.org/10.1007/s10459-016-9698-1

van der Lans, R., Pieters, R., & Wedel, M. (2008). Eye-Movement Analysis of Search Effectiveness. *Journal of the American Statistical Association*, *103*(482), 452–461. https://doi.org/10.1198/016214507000000437

van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Kramer, J., Warmuth, E., Heekeren, H. R., & Wartenburger, I. (2010). Resource allocation and fluid intelligence: Insights from pupillometry. *Psychophysiology*, *47*(1), 158–169.

van Meel-Jansen, A. (2006). The Magical Number Five in Art Appreciation. *Empirical Studies of the Arts*, *24*(1), 107–118. https://doi.org/10.2190/763E-H2QN-19Y0-A84N

van Merrienboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, *17*(2), 147–177.

Vansteenkiste, P., Cardon, G., Philippaerts, R., & Lenoir, M. (2015). Measuring dwell time percentage from head-mounted eye-tracking data--comparison of a frame-by-frame and a fixation-by-fixation analysis. *Ergonomics*, *58*(5), 712–721. https://doi.org/10.1080/00140139.2014.990524

Vermeersch, L., & Vandenbroucke, A. (2015). Kids, take a look at this! Visual Literacy Skills in the School Curriculum. *Journal of Visual Literacy*, *34*(1), 106–130. https://doi.org/10.1080/23796529.2015.11674725

Vermunt, J., K., & Magidson, J. (2004). Latent class analysis. In M. Lewis-Beck, A. E. Bryman, & T. F. Liao (Eds.), *The SAGE encyclopedia of social science research methods* (Vol. 2, pp. 549–553). Sage.

Viola, I., Chen, M., & Isenberg, T. (2020). Visual Abstraction. In M. Chen, H. Hauser, P. Rheingans, & G. Scheuermann (Eds.), *Foundations of Data Visualization* (pp. 15–37). Springer International Publishing. https://doi.org/10.1007/978-3-030-34444-3_2

Visser, I. (2011). Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. *Journal of Mathematical Psychology*, *55*(6), 403–415. https://doi.org/10.1016/j.jmp.2011.08.002

Visser, I., Raijmakers, M. E., & Molenaar, P. C. (2002). Fitting hidden Markov models to psychological data. *Scientific Programming*, *10*(3), 185–199. https://content.iospress.com/articles/scientific-programming/spr00112

Visser, I., & Speekenbrink, M. (2010). depmixS4 : An R Package for Hidden Markov Models. *Journal of Statistical Software*, *36*(7). https://doi.org/10.18637/jss.v036.i07

Vogt, S., & Magnussen, S. (2007). Expertise in pictorial perception: Eye-movement patterns and visual memory in artists and laymen. *Perception*, *36*(1), 91–100. https://doi.org/10.1068/p5262

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wagner, E. (2017). A Tool for Art Educators: The Common European Framework of Reference for Visual Literacy. In R. Mateus-Berr & L. Reitstätter (Eds.), *Edition Angewandte. Art & Design Education in Times of Change: Conversations Across Cultures* (pp. 115–120). De Gruyter. https://doi.org/10.1515/9783110528329-019

Wagner, E., & Schönau, D. (Eds.). (2016). *Common European Framework of Reference for Visual Literacy-Prototype*. Waxmann.

Wedel, M., & Pieters, R. (2017). A Review of Eye-Tracking Research in Marketing. In N. K. Malhotra (Ed.), *Review of Marketing Research* (pp. 123–147). Routledge. https://doi.org/10.4324/9781351550932-5

Weiß-Wittstadt, S. (2022). *Die Bedeutung sozialer Herkunft für Unterrichtswahrnehmung und Motivation im Kunstunterricht* (1. Auflage). *Empirische Erziehungswissenschaft: Vol. 80*. Waxmann.

Wiedermann, W., Frick, U., & Merkle, E. C. (2021). Detecting Heterogeneity of Intervention Effects in Comparative Judgments. *Prevention Science : The Official Journal of the Society for Prevention Research.* Advance online publication. https://doi.org/10.1007/s11121-021-01212-z

Wiedermann, W., Niggli, J., & Frick, U. (2014). The Lemming-effect: Harm perception of psychotropic substances among music festival visitors. *Health, Risk & Society*, *16*(4), 323–338. https://doi.org/10.1080/13698575.2014.930817

Winston, A. S., & Cupchik, G. C. (1992). The Evaluation of High Art and Popular Art By Naive and Experienced Viewers. *Visual Arts Research*, *18*(1), 1–14. http://www.jstor.org/stable/20715763

Winter, M., Pryss, R., Probst, T., & Reichert, M. (2021). Applying Eye Movement Modeling Examples to Guide Novices' Attention in the Comprehension of Process Models. *Brain Sciences*, *11*(1), 72. https://doi.org/10.3390/brainsci11010072

Witkin, R. W. (1983). The Psychology of Abstraction and the Visual Arts. *Leonardo, 16*(3), 200. https://doi.org/10.2307/1574914

Wolfe, J. M. (2010). Visual search. *Current Biology : CB, 20*(8), R346-9. https://doi.org/10.1016/j.cub.2010.02.016

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology. Human Perception and Performance*, *29*(2), 483–502. https://doi.org/10.1037/0096-1523.29.2.483

Wood, B. P. (1999). Visual expertise. *Radiology*, *211*(1), 1–3. https://doi.org/10.1148/radiology.211.1.r99ap431

Yang, J., Yan, F.-F., Chen, L., Xi, J., Fan, S., Zhang, P., Lu, Z.-L., & Huang, C.-B. (2020). General learning ability in perceptual learning. *Proceedings of the National Academy of Sciences*, *117*(32), 19092–19100. https://doi.org/10.1073/pnas.2002903117

Yarbus, A. L. (1967a). *Eye Movements and Vision*. Springer.

Yarbus, A. L. (1967b). Eye Movements During Perception of Complex Objects. In *Eye Movements and Vision* (pp. 171–211). Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-5379-7_8

Ylitalo, A.-K., Särkkä, A., & Guttorp, P. (2016). What we look at in paintings: A comparison between experienced and inexperienced art viewers. *The Annals of Applied Statistics*, *10*(2), 549–574. https://doi.org/10.1214/16-AOAS921

Yu, S.-Z. (2010). Hidden semi-Markov models. *Artificial Intelligence*, *174*(2), 215–243. https://doi.org/10.1016/j.artint.2009.11.011

Zangemeister, W. (1995). Evidence for a global scanpath strategy in viewing abstract compared with realistic images. *Neuropsychologia*, *33*(8), 1009–1025. https://doi.org/10.1016/0028-3932(95)00014-T

Zeileis, A., Hothorn, T., & Hornik, K. (2008). Model-Based Recursive Partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492–514. https://doi.org/10.1198/106186008X319331

Zemblys, R., Niehorster, D. C., Komogortsev, O., & Holmqvist, K. (2018). Using machine learning to detect events in eye-tracking data. *Behavior Research Methods*, *50*(1), 160–181. https://doi.org/10.3758/s13428-017-0860-3

Zhu, Y., Yan, Y., & Komogortsev, O. (2020). Hierarchical HMM for Eye Movement Classification. In (pp. 544–554). Springer, Cham. https://doi.org/10.1007/978-3-030-66415-2_35

Zimoch, M., Mohring, T., Pryss, R., Probst, T., Schlee, W., & Reichert, M. (2017). Using Insights from Cognitive Neuroscience to Investigate the Effects of Event-Driven Process Chains on Process Model Comprehension. In *International Conference on Business Process Management* (pp. 446–459). https://doi.org/10.1007/978-3-319-74030-0_35

Zimoch, M., Pryss, R., Layher, G., Neumann, H., Probst, T., Schlee, W., & Reichert, M. (2018). Utilizing the Capabilities Offered by Eye-Tracking to Foster Novices' Comprehension of Business Process Models, Cham.

Zimoch, M., Pryss, R., Schobel, J., & Reichert, M. (2017). Eye tracking experiments on process model comprehension: lessons learned. In J. Gulden (Ed.), *Enterprise, Business-Process and Information Systems Modeling* (pp. 153–168). Springer, Cham.

APPENDIX A

<table>
<tr><th>Eye-Tracking experiment</th><th>Specification Details</th></tr>
<tr><td>Eye-Tracker</td><td>SMI eye-tracking glasses were used (SMI ETG 2w Analysis Pro)
Video-based corneal reflection eye tracker</td></tr>
<tr><td>Eye-Tracking Software</td><td>SMI iView ETG for recording and calibration
SMI BeGaze for Semantic Gaze Mapping and statistical analysis</td></tr>
<tr><td>Video Stimuli</td><td>Video from SMI eye-tracking Glasses
Sampling Rate: 60 FPS
Calibration Area: 1290x980 pixels</td></tr>
<tr><td>Event Detection</td><td>SMI Event Detection (Velocity-based binocular event detection with saccades as primary events (Engbert et al., 2016)). The default eye movement parameters from SMI BeGaze™ version 3.7 were used. The following is taken directly from the BeGaze™ Manual:
The SMI ETG Event Detection algorithm pipeline:
<table>
<tr><th>Step</th><th>Description</th></tr>
<tr><td rowspan="3">1. Pre-processing</td><td>1.1 Convert the POR (Point of Regard) pixel values to degrees.</td></tr>
<tr><td>1.2 Compute velocity and acceleration of the POR (in degrees).</td></tr>
<tr><td>1.3 Compute velocity skewness (here defined as the ratio of velocity mean to velocity median over a 5-sample window).</td></tr>
<tr><td>2. Noise Detection</td><td>Identify single-sample spikes in the POR and remove them by interpolation.</td></tr>
<tr><td>3. Blink Detection</td><td>Identify Blinks based on pupil confidence (minimum duration of a blink event is 3 samples).</td></tr>
<tr><td rowspan="3">4. Saccade Detection</td><td>4.1 Detect midpoints of saccade candidates by searching for samples, which have either:
- POR velocity values above the threshold α_{def}, or
- POR velocity values above α_{min} and skewness above β.</td></tr>
<tr><td>4.2 Find beginnings and ends of saccade candidates by searching for local maxima in absolute POR acceleration values.</td></tr>
<tr><td>4.3 Accept saccade candidates as saccades if the detections for left and right eye are consistent.</td></tr>
<tr><td>5. Visual Intake Detection</td><td>Mark all the remaining samples as Visual Intake.</td></tr>
<tr><td>6. Post-processing</td><td>6.1 Remove saccade events smaller than γ in amplitude, or only one sample in duration, by interpolating with neighbors.</td></tr>
</table></td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><td rowspan="3"></td><td colspan="2">6.2 Mark Visual Intake events shorter than 50ms as “Undefined”.</td></tr>
<tr><td colspan="2">6.3 If Undefined event occurs immediately after saccade, merge Undefined with saccade.</td></tr>
<tr><td colspan="2">6.4 If Undefined event occurs immediately after blink, merge Undefined with Blink.</td></tr>
<tr><td>Threshold Name</td><td>Value</td><td>Units</td></tr>
<tr><td>α_{def}</td><td>100</td><td>°/s</td></tr>
<tr><td>α_{min}</td><td>8</td><td>°/s</td></tr>
<tr><td>β</td><td>5</td><td></td></tr>
<tr><td>γ</td><td>0.5</td><td>°</td></tr>
</table>
</td></tr>
<tr><td>Data collection and recording location</td><td>Subjects were seated 50-80 cm away from the tablet screen. Each session started with a 3-point calibration following the standard procedures for SMI iView™. The default eye movement parameters from SMI BeGaze™ version 3.7 were used. To collect data from Visual Literacy experts as well as novices multiple locations were allocated for recording sessions. These included a laboratory room at the HSD University of Applied Sciences in Cologne, Ulm University and a seminar room at the Academy of Fine Arts in Munich.
Only the experimenter (MT) and the participant were present during the recording session. The rooms were not soundproof, the luminance of the recording sessions was not controlled for and therefore might differ slightly.
The eye-tracking glasses were connected to a laptop PC where the recorded video and video footage of the eyes were displayed and monitored by the experimenter.
The stimuli were presented on Android A6 Tablets with 10.1 inch screen size. All test items were programmed specifically for the assessment tool (Andrews et al., 2018). The process models were created in BPMN 2.0 (OMG, 2011 OMG Specification, Object Management Group).
The equipment and procedure were explained to the participant before they put on the eye-tracking glasses.

All participants were given the identical instruction on the tablet screen:
“In the following, different processes are presented in the form of process models. A process model visualizes the sequence of events and decisions. Try to understand the process in the process model and select all correct statements (multiple statements can be correct).”</td></tr>
</table>

	[In German: “Im Folgenden werden verschiedene Abläufe in Prozessmodellen präsentiert. Ein Prozessmodell stellt Abläufe von Ereignissen und Entscheidungen visuell dar. Versuche den Ablauf in den Prozessmodellen nachzuvollziehen und wähle alle richtigen Aussagen aus (auch mehrere Aussagen können richtig sein).”] During recording the experimenter sat behind or next to the participant at a separate table. All participants gave written informed consent. Andrews, K., Zimoch, M., Reichert, M., Tallon, M., Frick, U., & Pryss, R. (2018). A Smart Mobile Assessment Tool for Collecting Data in Large-Scale Educational Studies. *Procedia computer science*, *134*, 67-74. Omg, O.M.G. (2011 OMG Specification, Object Management Group.). *Business Process Model and Notation (BPMN) Version 2.0.* [Online]. https://www.omg.org/spec/BPMN/2.0/: OMG Group. [Accessed November 2018].
Calibration	3-point calibration before each of the 5 tasks on Visual literacy including the Process Model Task • Black dots on white background placed in the top and in the lower corners of the tablet screen. • Participants were told by the experimenter to look at each of the three dots in consecutive order. • Re-calibration was done if the participants gaze was not located at the correct points during validation phase. Re-calibration was also conducted when the signal was too low during the experiment (indicated by the SMI ETG warning message). • Validity of calibration was also monitored throughout the recording session (gaze overlaid video was displayed on the experimenter’s notebook), video recordings of the eye was visible during calibration) • Participants were instructed to look at a fixation cross that was displayed between each model for 2 seconds.
Participation and Excluded Trials	Participants in Sample II were enrolled as experts in Visual Literacy, if they were members of the European Network of Visual Literacy (ENViL) or working in professions requiring a high visual competence (photographer, gallerist, art educator, art designer, art students, or self-employed artists). Novices in Visual Literacy were adults from the clerical and academic staff of various educational settings declaring themselves as not overwhelmingly talented or familiar with arts and visual design.

<table>
<tr><td></td><td>Participants were asked if they wore glasses or contact lenses. If they wore glasses, they were offered SMI corrective lenses and were able to take part in the experiment with corrected-to-normal vision.

Participants were excluded if they did not fulfil the inclusion criterion vision, in cases where they had poor near acuity (for example could not read the instruction texts on tablets even with SMI corrective lenses (+2.5 diopter lenses)) (n=2)

Screened for eligibility: 41 participants
Non eligible according to our predefined inclusion and exclusion criteria: 5 (2 severe limitation in visual acuity, 3 were excluded from analysis because the data quality was too low).
Analyzed sample: 36 participants satisfied all inclusion and exclusion criteria</td></tr>
<tr><td>Quality Threshold for analysis</td><td>BeGaze SMI Tracking Ratio needed to be > 95% (n=3 participants excluded)
Tracking ratio is defined as the number of non-zero gaze positions divided by sampling frequency multiplied by run duration expressed in percent. This threshold was used because in pretests with the eye-tracking glasses, visual inspection of raw data showed frame loss in participants had to be below 5%.

The mean Tracking Ratio for the analyzed sample was 97.76 %</td></tr>
<tr><td>Reference Image:</td><td>3 reference images (one for each model)
Size: 1920x1200 pixels (same as image on tablet screens)</td></tr>
<tr><td>AOI related metrics</td><td>
<table>
<tr><td>AOI Name</td><td>AOI Size [px]</td><td>AOI Coverage [%] on reference image</td></tr>
<tr><td>PM 1
PM 2
PM 3</td><td>1033444
1033444
1033444</td><td>44.9
44.9
44.9</td></tr>
<tr><td>Relevant model parts*</td><td>294400</td><td>12.8</td></tr>
<tr><td>Irrelevant model parts*</td><td>739044</td><td>32.1</td></tr>
<tr><td>Statements 1-4</td><td>124493 (x4)</td><td>5.4 (x4)</td></tr>
<tr><td>White Space PM1
White Space PM2
White Space PM3</td><td>773420
773420
773420</td><td>33.6
33.6
33.6</td></tr>
</table>

*only for Process Model 2</td></tr>
</table>

APPENDIX B

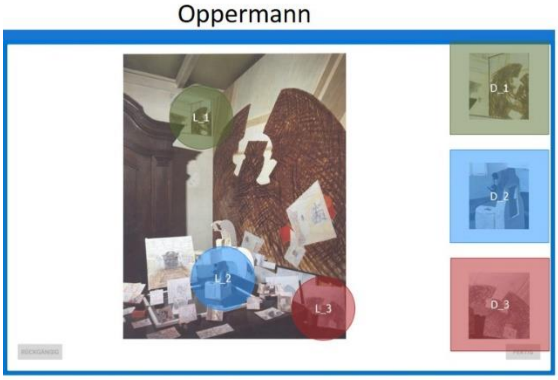
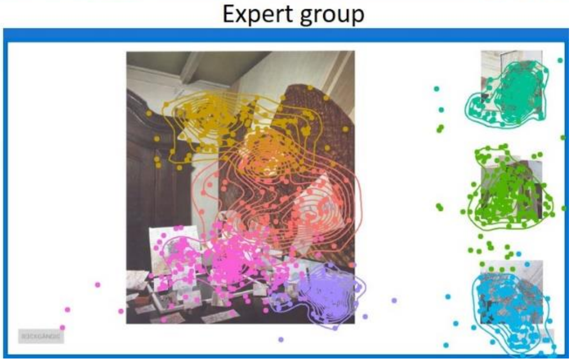
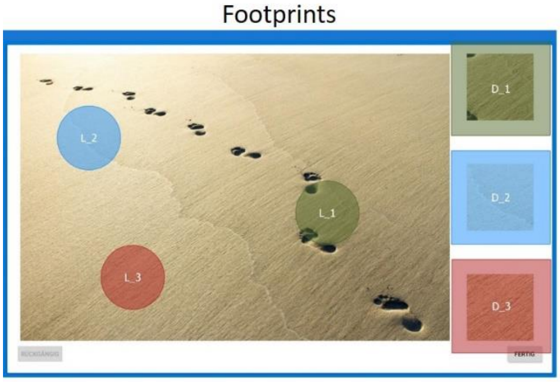
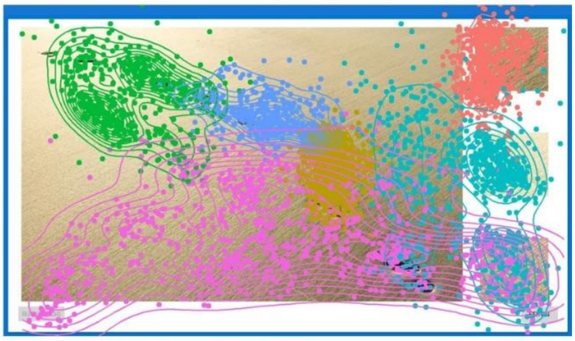
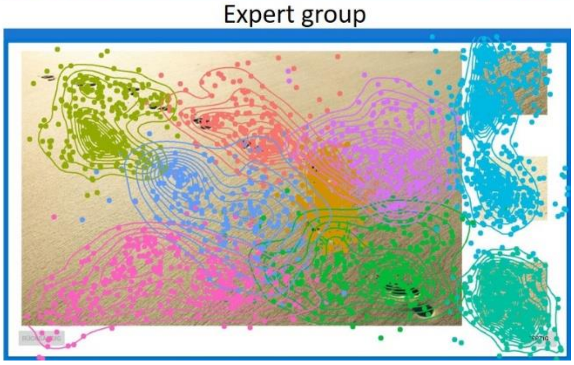
Image title	HMM Density map
Oppermann 	Non-expert group  Expert group 
Footprints 	Non-expert group  Expert group 

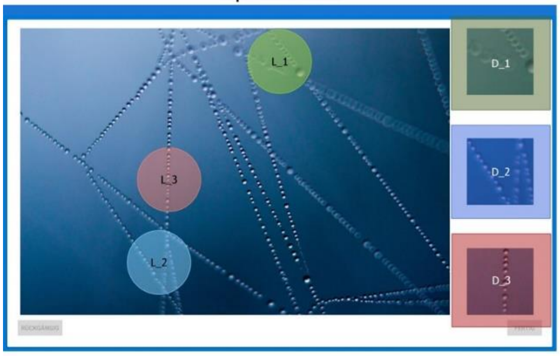
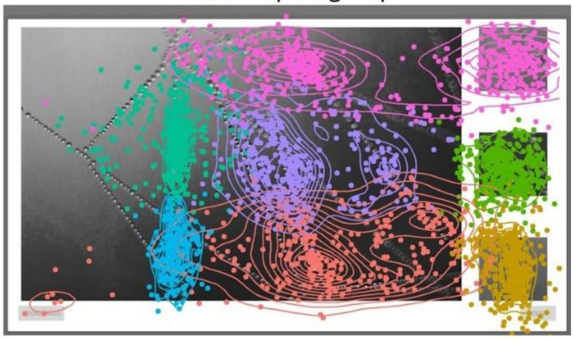
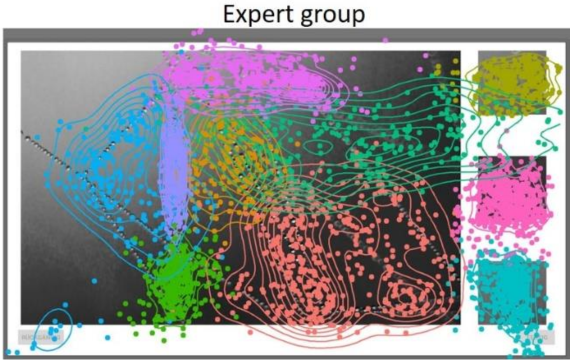
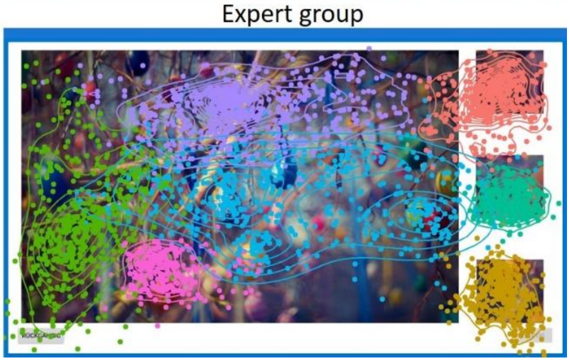
Image title	HMM Density map
Spider Net 	Non-expert group  Expert group 
Easter 	Non-expert group  Expert group 

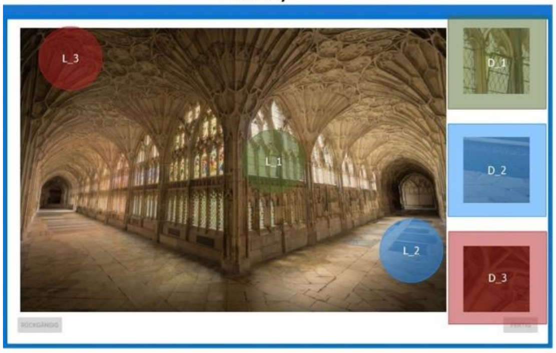
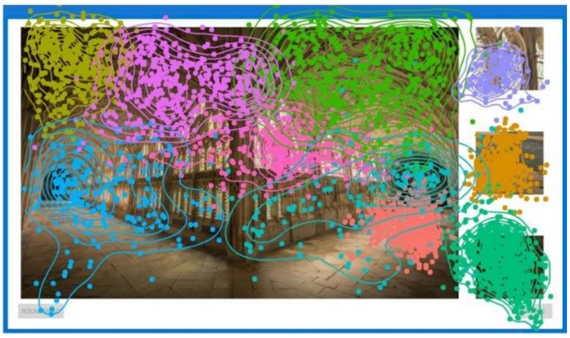
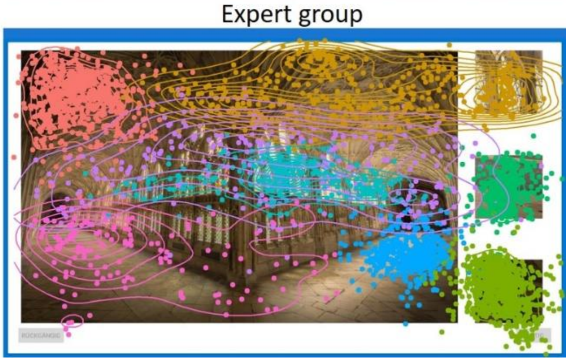
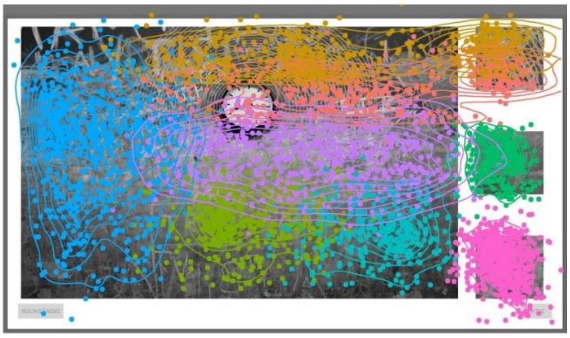
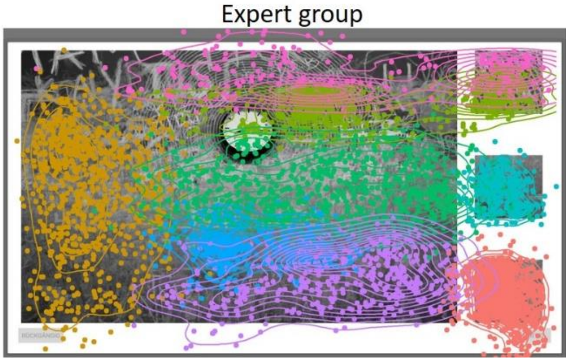
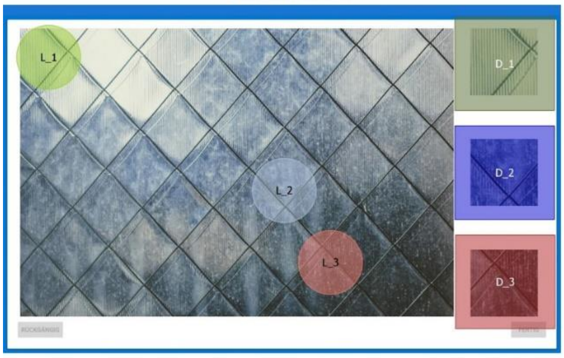
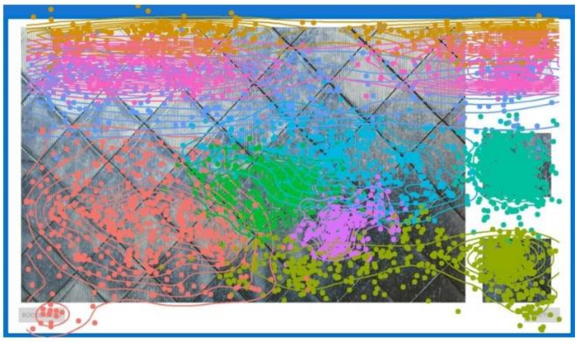
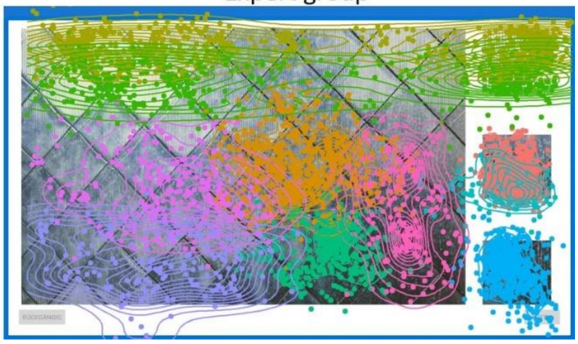
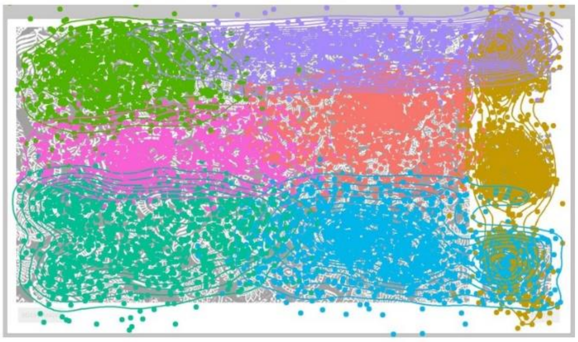
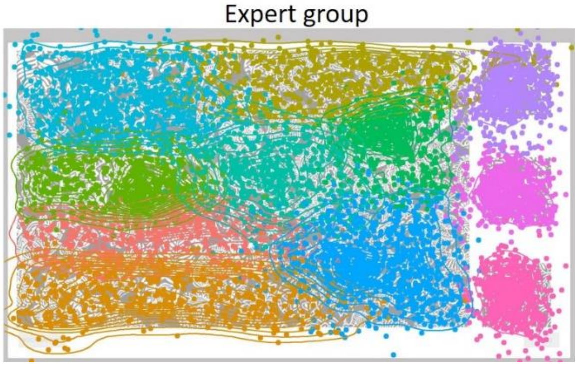
Image title	HMM Density map
Abbey 	Non-expert group  Expert group 
Clock & Graffiti 	Non-expert group  Expert group 

Image title	HMM Density map
Glasshouse 	Non-expert group  Expert group 
Linocut Pattern 	Non-expert group  Expert group 

APPENDIX C

DAS TEAM DES BKKB-VERBUNDPROJEKTES

Prof. Dr. Katrin Rakoczy
Projektleitung

Prof. Dr. Ulrich Frick
Projektleitung

Susanne Weiß M.A.
Wissenschaftliche
Mitarbeiterin

Miles Tallon M.Sc.
Wissenschaftlicher
Mitarbeiter

WIR HABEN IHR INTERESSE GEWECKT?

Wenn Sie an dem Verbundprojekt „Bildkompetenz in der kulturellen Bildung (BKKB). Was ist und wie fördert man Bildkompetenz“ teilnehmen oder gern weitere Informationen zur Feldstudie des Verbundprojekts erhalten möchten, dann kontaktieren Sie uns:

Miles Tallon M.Sc.
Wissenschaftlicher Mitarbeiter
Hochschule Döpfer (HSD)
Waidmarkt 3 und 9
50676 Köln
Tel.: 0221 - 126 125 27
m.tallon@hs-doepfer.de

Susanne Weiß M.A.
Wissenschaftliche Mitarbeiterin
Deutsches Institut für Internationale Pädagogische Forschung (DIPF)
Schloßstrasse 29
60486 Frankfurt
Tel: 069 - 247 08-165
weiss2@dipf.de

BILDKOMPETENZ IN DER KULTURELLEN BILDUNG UNTERRICHT

BILDKOMPETENZ IN DER KULTURELLEN BILDUNG ASSESSMENT

HSD Hochschule Döpfer University of applied sciences

DIPF

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

BILDKOMPETENZ IN DER KULTURELLEN BILDUNG

Was ist und wie fördert man Bildkompetenz?

Wie Sie bei der Feldstudie des Verbundprojektes im Schuljahr 2017/18 teilnehmen können.

WAS IST DAS BKKB-VERBUNDPROJEKT?

Ziel des Forschungsprojekts ist es, in verschiedenen Teilstudien die Qualität von Kunstunterricht und seine Wirkungen zu untersuchen. Wir wollen von und mit Ihnen lernen, was guten und für die Schülerinnen und Schüler interessanten Kunstunterricht ausmacht und Folgendes untersuchen:

1. Wie kann Bildkompetenz als zentrales Konstrukt in der kulturellen Bildung empirisch messbar gemacht werden?
2. Wie kann schulischer Kunstunterricht gezielt dazu beitragen, Motivation und Bildkompetenz zu fördern?
3. Wie prägen Merkmale sozialer Herkunft (sozialer Status der Eltern, familiäre Kapitalzusammensetzung, kulturell-ästhetische Praxis im Elternhaus) die Wahrnehmung des Kunstunterrichts durch die Lernenden und letztendlich deren Lernerfolg?

Aus den Ergebnissen verschiedener Teilstudien sollen fundierte Erkenntnisse zur qualitätsvollen Gestaltung von Kunstunterricht und der gezielten Förderung von Motivation und Bildkompetenz im Rahmen von Kunstunterricht abgeleitet werden. Diese werden nach Ende der Projektlaufzeit an die teilnehmenden Kunstlehrkräfte zurückgemeldet und in die Schulpraxis des Kunstunterrichts eingebracht.

Dazu sind wir auf engagierte Kunstpädagoginnen und Kunstpädagogen, die einen Beitrag zur Erforschung von Lehr-Lernprozessen im Kunstunterricht und zur gezielten Förderung von Bildkompetenz im Rahmen schulischen Kunstunterrichts leisten wollen, als Kooperationspartner angewiesen.

WIR BRAUCHEN IHRE UNTERSTÜTZUNG!

Wir sind auf der Suche nach engagierten Kunstlehrkräften der Sekundarstufe I:

- Für das Schuljahr 2017/18 suchen wir insgesamt 40 Lehrkräfte in Nordrhein-Westfalen, Hessen und Bayern
- Sie sollten eine Kunstklasse der Jahrgangsstufe 9 unterrichten

WIE LÄUFT DIE STUDIE AB?

Das Projektteam kommt an Ihre Schule und führt eine 90-minütige Erhebung während der regulären Unterrichtszeit durch.

- Befragung der Schülerinnen und Schüler
  - Dauer: 45 Minuten
  - Form: Elektronisch via Tablet
  - Inhalt: Unterrichtswahrnehmungen, Motivation, familiäre Hintergrundmerkmale
- Testung der Schülerinnen und Schüler
  - Dauer: 45 Minuten
  - Form: Elektronisch via Tablet
  - Inhalt: Bildkompetenz
- Befragung der Kunstlehrkraft
  - Dauer: ca. 30 Min. (während der Schülerbefragung)
  - Form: Papierversion
  - Inhalt: Unterrichtsgestaltung, Hintergrundmerkmale
- Befragung der Eltern
  - Dauer: ca. 30 Minuten
  - Form: Papierversion (wird den Schülerinnen und Schülern mitgegeben)
  - Inhalt: Familiäre Hintergrundmerkmale
- Eyetracking-Studie mit einzelnen Schülerinnen und Schülern

Selbstverständlich...

...ist die Teilnahme an der Studie für Sie und Ihre Schülerinnen und Schüler freiwillig.

...werden alle im Rahmen der Studie gesammelten Daten, Tests und Fragebögen streng vertraulich behandelt und nur zum Zwecke der wissenschaftlichen Erforschung verwendet.

WAS BRINGT IHNEN DIE TEILNAHME?

- Individuelle Rückmeldung zu Ihrem Kunstunterricht aus Sicht der Schülerinnen und Schüler
- Rückmeldung über die Ergebnisse der Teilstudien des Verbundprojekts
- 100 € für die Klassenkasse für jede teilnehmende Kunstklasse oder für den Förderverein Ihrer Schule

[Leaflets used for the recruitment]