Learning from Teacher's Eye Movement: Expertise, Subject Matter and Video Modeling

by

Yizhen Huang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Combined Program in Education and Psychology)
in The University of Michigan
2018

Doctoral Committee:

Professor Kevin F. Miller, Chair
Professor Kai Schnabel Cortina
Professor Richard D. Gonzalez
Professor Priti R. Shah

Yizhen Huang
yizhenh@umich.edu
ORCID iD: 0000-0002-7041-1927

For all the teachers that shape the future.

ACKNOWLEDGEMENTS

This work would not be possible without the guidance from my committee: Prof. Kevin F. Miller, Prof. Kai Schnabel Cortina, Prof. Richard D. Gonzalez and Prof. Priti R. Shah. Their incomparable devotion to studying the human mind inspired me as a researcher, and their impeccable characters have shaped my life as a person. A special thank you goes to Prof. Kevin F. Miller, who has became a teacher, a guide, a friend and also a family member to me, just like the Chinese saying: “Once a mentor, always a father”.

I also want to thank you for all the participating teachers in the studies. They have built, or are on the way to building a brighter future for all their students. I saw passion, ingenuity, and unconditional love in their everyday classrooms.

Thank you to all the colleagues that I have honor to work with throughout my graduate years: my labmates in VizLab and my colleagues in the Digital Education and Innovation Lab (DEIL). Their creative and original thoughts have always inspired me to explore new territories.

And finally, a wholehearted thank you to my family and friends who supported me through thick and thin. They are my safe harbor even we are thousands of miles away.

TABLE OF CONTENTS

CHAPTER

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

How teachers' eye movements can be used to understand and improve education is the central focus of the present paper. Three empirical studies were carried out to understand the nature of teachers' eye movements in natural settings and how they might be used to promote learning. The studies explored 1) the relationship between teacher expertise and eye movement in the course of teaching, 2) how individual differences and the demands of different subjects affect teachers' eye movement during literacy and mathematics instruction, 3) whether including an expert's eye movement and hand information in instructional videos can promote learning. Each study looked at the nature and use of teacher eye movements from a different angle but collectively converge on contributions to answering the question: what can we learn from teachers' eye movements? The paper also contains an independent methodology chapter dedicated to reviewing and comparing methods of representing eye movements in order to determine a suitable statistical procedure for representing the richness of current and similar eye tracking data.

Results show that there are considerable differences between expert and novice teachers' eye movement in a real teaching situation, replicating similar patterns revealed by past studies on expertise and gaze behavior in athletics and other fields. This paper also identified the mix of person-specific and subject-specific eye movement patterns that occur when the same teacher teaches different topics to the same children. The final study reports evidence that eye movement can be useful in teaching; by showing increased learning when learners saw an expert model's eye movement in a video modeling example. The implications of these studies regarding teacher education and instruction are discussed.

CHAPTER I

Introduction

Vision is arguably our principal perceptual modality. Human vision serves the critical function of building a bridge between the world and our mind. The incessant movement of our eyes not only gathers information from our surroundings, but it can also reveal something about our own knowledge and internal states. Knowing where and how we look at the world can open a unique gateway into otherwise hidden mental activities. Thus eye tracking technology that captures human eye movement affords us a great opportunity for studying the human mind.

Teachers, students and educational material form an instructional triangle that enables the transmission of knowledge from generation to generation (*Ball*, 2000; *Lampert and Ball*, 1998). For as long as human beings have existed, we have passed on knowledge and skill through these intimate interactions between teachers and students. Although teaching and learning are observable and assessable as overt phenomena, what we really care about are inherently covert cognitive processes that are difficult to access. Tracking people's eye movements during teaching and learning has great potential for revealing hidden mental activity in a non-intrusive and non-disruptive manner.

Most existing eye tracking research on learning has focused on students' looking during very constrained circumstances (typically sitting in front of a computer). We know relatively little about teacher looking during teaching, or about student looking in realistic natural situations. For example, from studies in high-level cognition, we already know that learners' eye movements serve as an embodied sensorimotor part of the cognitive process that are invoked by linguistic transmission or spatio-temporal arrangement during problem solving (*Spivey and Dale*, 2011, p. 552). Compared to our relatively extensive knowledge about learners' eye movements, the teacher's perspective is often hidden from view.

Is the invisibility of the teacher's viewpoint an indicator that teaching is less relevant in studying human cognition? Quite the contrary; teaching is one of the most cognitively demanding tasks. Teachers have to navigate through the classroom, operate classroom equipment, monitor students' behavior and attention, manage student comprehension and engagement, all at the same time they endeavor to deliver a coherent, even interactive, lesson. The irreducible complexity of teaching requires enormous attentional resources and skillful attention allocation. Given the complexity and interactive nature of teaching, measures that are unintrusive, dynamic, and capable of capturing human attention —such as eye movement—are an attractive choice for learning about teaching.

Objectives The primary motivation of the current paper is to learn what we can uncover from studying teachers' eye movements. The current endeavor is initiated with the belief that capturing teachers' eye movements can deepen our understanding about teaching by addressing the following questions: 1) how do teachers distribute attention in the course of instruction, 2) how can we represent the regularities in how teachers move their eyes, 3) what is the influence of expertise on teacher eye movements, 4) to what extent do eye movement patterns vary when the same teacher is teaching different subject areas. and 5) can we apply expert's eye movement in instructional videos?

The secondary motivation of this paper is to identify the unique challenges of analyzing eye movement data. These data comprise a set that is inherently rich and complex, containing both spatial and temporal information. We will attempt to describe a statistical work flow that fully utilizes the richness and complexity of eye movement data, particularly for purposes of statistical inference.

Structure of the Dissertation To support the stated objectives, this paper will present three empirical studies. Each study looks at the nature and use of teacher eye movements from a different angle but collectively converge on contributions to answering the question: what can we learn from teachers' eye movements? The paper also contains an independent methodology chapter dedicated to reviewing, contrasting and determine a suitable statistical procedure for the representing the richness of current and similar eye tracking data.

The paper begins with a review of the types of eye movement and the current theoretical landscape of factors affecting eye movement. Then the Methodology chapter concentrates on finding a suitable statistical analysis method by comparing existing approaches to eye movement data. The unique challenges of free-moving, free-viewing mobile eye tracking data will also be discussed.

These methods will then be applied to two mobile eye tracking studies of teacher looking in the course of instruction. Study 1 compares novice and experienced teachers teaching the same topics to the same children, asking how experienced teachers' eye movements differ from those of novices. Study 2 explores the effects of topic of instruction, taking advantage of the fact that American elementary school teachers teach multiple subjects to the same students. We will compare the same teachers teaching literacy and mathematics to see the extent to which teachers show consistent patterns of looking across topics, or the extent to which different subjects lead to different looking patterns. And finally, the benefit of showing expert eye movement in instructional videos will be explored in Study 3.

Finally the implications of these studies regarding teacher education and instruction will be covered in the Discussion chapter.

1.1 A Brief Review of Eye Movement Research

Overview of Literature Review The literature review section starts with a short introduction about the basic mechanisms and typical types of eye movements. Models of looking can be distinguished based on the emphasis they place on bottom-up versus top-down processes. Several of the most influential models for explaining and predicting fixation allocation and eye movement sequences will be described and compared. The strengths and limitations of those models will also be discussed in preparation for the Methodology chapter.

Typical Eye Movement Events Human vision is a peculiar system. Contrary to many people's intuition and subjective experience that eyes take in visual information smoothly and continuously, the eyes literally move in fits and starts—fixations and then jumps to new locations (*Findlay and Gilchrist*, 2008; *Richardson and Spivey*, 2004). This saccade-fixate-saccade mechanism evolved as a solution to the problem that, despite the abundance of information our visual system takes in at any given time, there is only a limited region at the rear surface of the eye called the *fovea* that has the sensitivity and capability to process that information in detail. Information quality decreases substantially away from the center of the gaze. *Bouma* (1970) showed that for a letter presented at different eccentricities from the fovea, the rate of accurately identifying the letter dropped from 100% to 50% at 3° and 10° respectively (*Gilchrist*, 2011, p. 85). To exploit this narrow optimal spot and ensure high acuity visual information, human eyes have to rapidly scan the scene and focus on new informative areas. Eye movement is thus crucial to direct fixations in the service of the ongoing perceptual, cognitive and behavioral activity (*Henderson*, 2011, p. 593).

When scanning and sampling visual scenes, the eyes never completely rest on a single spot. Instead, fixations always contain involuntary miniature eye movements such as *tremor*, *drift* and *microsacaades* (see 1.1) (*Duchowski*, 2007b; *Holmqvist et al.*, 2011a; *Gilchrist*, 2011). But fixations can still be defined by the relative stillness of the eyeball for a short period of time. The fact that the eyes stabilize over a spot is generally accepted as a sign of attention, with the assumption that useful visual information is being gathered when the eye is relatively stationary (*Wade and Tatler*, 2011, p. 29), though other possible explanation also exist (*Holmqvist et al.*, 2011a, p. 22). Fixation is also claimed to be tightly tied into deeper cognitive processes such as detail perception, pattern recognition, memory encoding and language processing (*Henderson and Hollingworth*, 1999; *Hollingworth et al.*, 2001; *Ballard et al.*, 1995; *Nelson and Loftus*, 1980; *Meyer and Lethaus*, 2004; *Tanenhaus et al.*, 2004; *Henderson*, 2011). Thus fixation provides us with an observable, unobtrusive and real-time behavioral index of the underlying cognitive processing (*Rayner*, 1998).

As noted already, the eyes are constantly moving even during fixations. Unlike many other physiological and behavioral measures, oculomotor events consist of both voluntary and involuntary movements. There are five basic types of normal eye movements: saccadic, smooth pursuit, vergence, vestibular, and physiological nystagmus (tremor) (*Duchowski*, 2007b; *Carpenter*, 1988; *Robinson*, 1968). Psychologists interested in cognition and perception have focused on only two of these kinds of eye movements: saccades and smooth pursuits, because they are the most closely connected to voluntary activity. Other forms of eye movements are generally studied in the context of human neurology and may not provide useful information about cognitive processes; thus they are outside the scope of this paper.

Saccades are fast, stereotyped and ballistic eye movements that precede as well as follow fixations (*Gilchrist*, 2011; *Duchowski*, 2007b). It is believed that the main function of saccades is to rapidly reposition the fovea and concentrate visual attention on a new region of interest (*Walls*, 1962).

Saccades are characterized by short time duration, high velocity and wide range of amplitudes. Saccadic movements are combinations of target-elicited reflexive movements and goal-oriented voluntary movements. When making saccades, the eyes start from an initial stable state (fixation) and then quickly accelerate and reach a peak velocity that's immediately followed by a rapid deceleration back to a stable state (*Gilchrist*, 2011, p. 86). Saccades are said to be stereotyped and ballistic. Stereotyped in the sense that a) particular movement patterns appear repeatedly; and b) different individuals tend to have different temporal profiles. Ballistic refers to the assumption that the destination of saccades is predetermined and once a saccade is initiated, the ocular-motor system will enter a period called *dead-time*, during which the saccade cannot be altered even if the target changes (*Gilchrist*, 2011; *Duchowski*, 2007b).

When both the head and scene are still, eye movements are mainly saccadic. But in a more naturalistic environment such as a classroom with both the subject and objects constantly moving, it is the combination of different kinds of eye movements acting together that delivers visual information and stability. Smooth pursuit movements are an indispensable element in this mixture. The principal objective of smooth pursuit is to match the eye velocity with the object velocity, in order to reduce distortion and maintain resolution acuity (*Barnes*, 2011, p. 115). In mobile eye tracking situations, where both the target and the observer may be moving, smooth pursuit movements are difficult to identify. They also seem likely to be a small part of classroom information processing, and so they were not used as an eye movement measure in the current paper.

Table 1.1: Summary of Typical Eye Movement Events

Event	Attributes	Duration (ms)	Amplitude	Velocity
Fixation	Eyes rest still on a target for a short period of time	100–600	Not applicable	Not applicable
Microsaccade	Quick eye movements that bring the eye back to target	10–30	10–40’	15–50° / s
Tremor (physiological nystagmus)	High frequency involuntary and eye movement	Not applicable	1’	20’ / s (peak)
Drift	Slow eye movement that take the eye away from the center of fixation	200–1000	1–60’	6–25’ / s
Saccade	Fast, stereotyped and ballistic eye movement between two targets	10–100	4–20°	30–500° / s
Smooth pursuit	Slower smooth eye movement that track target continuously	Not applicable	Not applicable	10–30° / s

[2]

Where Do We Look? Real-world scenes are rich in visual stimuli, yet both our attention and visual processing capability are limited resources. Within a complex visual environment, the scanning process is neither uniform nor complete—only a small portion of elements have the privilege of receiving fixations (*Tatler et al.*, 2005). How we solve the puzzle of how and where to distribute visual focus greatly shapes our perception and understanding of the world.

Two broad frameworks have been proposed to determine where fixations will be directed. One is the bottom-up process driven by physical features of the visual stimulus (e.g., *Itti and Koch*, 2000), the other one is the top-down process generated by knowledge structure and experience (e.g., *Yarbus*, 1967b; *DeAngelus and Pelz*, 2009; *Borji et al.*, 2015; *Boisvert and Bruce*, 2016).

In the context of scene perception, the bottom-up mechanism refers to exogenous attention driven by low-level features of visual stimuli such as movement, luminance, color, contrast and edges (*Boisvert and Bruce*, 2016). The bottom-up process is almost instantaneous, independent of task and often obligatory (*Kowler*, 2011). Top-down mechanisms, on the other hand, are slower and goal-directed, controlled by endogenous aspects of attention, and influenced by experiences, expectations, intentions and task directives, etc. (*Parkhurst et al.*, 2002).

The extent to which each mechanism determines and influences eye movement has been under debate for several decades, but we have seen a trend of integration in recent years. Next I'll summarize the major research findings from each perspective.

Bottom-up Eye Movement Control Based on the assumption that fixation allocation is primarily controlled by visual features, the obvious first question for researchers is what distinguishes the areas that attract fixation from those that do not. Generally speaking, researchers have approached this question with two methods: either they a) directly manipulate visual features and examine the outcome, or b) observe eye movement behavior without manipulation and establish correlational relationships between visual features and eye movements, then further compare the prediction of a predetermined model with observed data.

Direct manipulation and comparison of visual features had been the default experimental method in the early days of bottom-up research. These studies had focused on the properties of visual stimuli in determining object detection, localization and recognition. *Treisman and Gelade* (1980) conducted a series of experiments (*Treisman*, 1982, 1983) using simple visual search tasks and discovered that targets with only one feature different from the background were processed in parallel across the entire search field (*pop-out* effect). For example, no matter where a green line is, it can be spotted against red lines with the same speed. Additionally, this process will change from parallel to a serial, self-terminating scan if the search target is defined by multiple features (such as color plus angle). *Julesz* (1984) identified a set of elementary visual features that can be processed in parallel. These features, termed textons, include: color, orientation of line segments, and certain shape parameters such as curvature (*Bergen and Julesz*, 1983; *Julesz*, 1984). *Koch and Ullman* (1985) expanded the list to include edges, direction of movement and disparity. Other features that are believed to automatically attract attention include abrupt onsets (*Yantis and Jonides*, 1984, 1996), the occurrence of a new stimulus (*Hillstrom and Yantis*, 1994) and unique features (*Treisman and Gelade*, 1980).

The early researchers also proposed a two-stage theory of human visual perception, divided into a pre-attentive mode and an attentive mode. First, all simple features are processed rapidly and in parallel over the entire visual field, and second, specialized focus is placed on particular locations, leading to recognition of objects (*Koch and Ullman*, 1985; *Treisman*, 1983; *Bergen and Julesz*, 1983). This theoretical construct built the basis for the dominant methodological model in the field—the saliency model.

This model started with *Koch and Ullman* (1985) proposing the concept of a *saliency map* to describe the conspicuity of low-level, elementary features of a location in the visual scene. The saliency map was defined as a "measure of conspicuity" derived from the local contrast of features such as luminance, color, and motion. According to the saliency model, the attention process starts with separating the visual input into multiple feature channels (color, luminance, orientation etc.). These simple feature representations are then optimized to detect local feature differences. Finally, a saliency map is calculated by aggregating the local feature differences across spatial scale and feature types. Thus a saliency map is a computational representation of the visual importance of stimuli at each location, and is based purely on stimulus features in the scene (*Koch and Ullman*, 1985). This map was believed to be computed at a very early visual processing stage, prior to the identification of objects. *Koch and Ullman* (1985) and subsequent researchers have linked the saliency map to the generation and distribution of attention reflected in fixations when perceiving a scene: the higher the computed salience values are, the more likely a given location would be fixated (e.g., *Peters et al.*, 2005; *Itti et al.*, 1998; *Itti and Koch*, 2001; *Itti*, 2005).

Although extensive research has been conducted in well-controlled experimental paradigms with simple stimuli and arbitrary tasks, how bottom-up mechanisms determine fixation allocation in complicated natural scenes has been examined much less extensively (*Derrick J. Parkhurst*, 2004). More recent work in this line of research has employed free-viewing rather than artificial tasks (e.g., visual search) to study the impact of different visual features. Many of these studies are built on large samples of fixations and have utilized computational modeling methods to study the relationship between the observed fixation locations and the saliency of those locations. For example, *Parkhurst et al.* (2002) recorded participants' eye movements while they free-viewed 300 natural and artificial scenes and examined the relationship between the fixation locations and saliency of the locations using the saliency model proposed by *Itti et al.* (1998). Their results were similar to those of early studies using simplified stimuli in indicating that attention can be guided by bottom-up mechanisms.

As prominent as it is, the saliency model has left out the influence of higher level cognitive factors including the viewer's knowledge, experience and identity, as well as the properties of the given visual task and strategies for completing it (*Bacon and Egeth*, 1994; *Ballard and Hayhoe*, 2009). This neglect is particularly problematic when visual perception is carried out in natural settings; studies utilizing natural viewing tasks have shown that the magnitude of bottom-up mechanisms' influence is largest for early fixations but gradually declines for subsequent fixations (*Parkhurst et al.*, 2002; *Derrick J. Parkhurst*, 2004). Multiple studies have also shown that when viewing static natural images, the predictive power of the saliency model only barely exceeds chance (*Schutz et al.*, 2011; *Tatler and Vincent*, 2009a; *Betz*, 2010). It appears that bottom-up mechanisms will operate automatically during a task only if attention is not deliberately allocated to a stimulus prior to new stimulus onset (*Theeuwes*, 1994; *Yantis and Jonides*, 1996; *Yantis and Egeth*, 1999). The lack of power of the saliency model in predicting task-dependent eye movements has led to the rise of top-down approaches.

Top-down Eye-movement Control The very first research in top-down eye movement control is also one of the pioneering studies in the whole field. *Yarbus* (1967a) showed with compelling examples that different tasks can induce different viewing patterns. In his study, viewers' eye movements were recorded while viewing a famous painting *They did not expect him* (by Ilya Repin) under different task instructions, for example, estimating the wealth of the family or the age of characters, guessing what the family had been doing, or memorizing the clothes worn by the characters, etc. Different gaze behavior was discovered for these different tasks, revealing both individual and task-specific eye movement patterns. *Yarbus* (1967a)'s results showed that the eyes are not necessarily drawn to overall salient areas, but instead, the areas that attract people's fixations are those that provide the most information for the given task (*DeAngelus and Pelz*, 2009). This was the first attempt to draw connections between eye movement patterns and high-level cognitive factors.

Following the work of Yarbus, studies that incorporate the role of task properties have shown that eye movements are deployed to extract very specific information needed by the ongoing task (*Jovancevic et al.*, 2006; *Ballard and Hayhoe*, 2009). For example, in a classic study about cricket batsmen, *Land and McLeod* (2000) found that batsmen would first fixate on the pitch release to get information on the ball's upcoming trajectory and then make a predictive saccade to the expected bouncing position. This gaze behavior can not be explained by visual salience since the landing point of a ball is essentially featureless.

On the other hand, subjects' personal characteristics such as gender (*Coutrot et al.*, 2016), age group (*French et al.*, 2017), expertise level (*Kübler et al.*, 2015; *Boccignone et al.*, 2014) and state of health (*Tseng et al.*, 2013), are also related to specific eye movement patterns. A more recent trend in this line of work is to reverse engineer the influence of top-down factors and task properties, proving that eye movements contain a wealth of information that reveals hidden patterns about both the viewer and the task (e.g., *Greene et al.*, 2012; *Kit and Sullivan*, 2016; *Tseng et al.*, 2013; *Kanan et al.*, 2014; *Haji-Abolhassani and Clark*, 2013, 2014).

Integration of Two Models Since scene features, viewer characteristics or task properties alone are insufficient to explain gaze behavior, an integration of approaches that acknowledges the joint influence of different factors has been proposed.

The early integrations can be categorized as incorporating either a weak or strong top-down hypothesis (*Betz*, 2010). The weak top-down hypothesis proclaims that top-down factors only influence eye movement through the modulation of the bottom-up system as feature weights are affected by top-down processes (*Itti and Koch*, 2001). In contrast, a strong top-down hypothesis proposes that top-down and bottom-up processes behave in independent fashion, and both processes directly affect the allocation of visual attention (*Ahissar and Hochstein*, 1997; *Betz*, 2010).

More recent integrations are less concerned about the relative importance of each process but rather consider visual processing as a unified process using both kinds of information. For example, *Ballard and Hayhoe* (2009) argues that in the visual perception of a scene during free-viewing, low-level features and the kind of vision involved in specific tasks are not necessarily distinguishable. They reasoned that “all vision can be conceptualized as a task of some kind.” Similarly, *Schutz et al.* (2011) proposed an integrative model that constitutes several interacting control loops: salience, object recognition, value and plans (*Fuster*, 2004). These elements work on different levels of processing, and jointly determines saccadic target selection.

With the development of more affordable and portable eye trackers, the discussion of bottom-up and top-down process will extend to more naturalistic viewing conditions involving complicated daily viewing tasks, where each factor can be better understood in natural settings. Under this framework, the task of teaching appears to be an excellent candidate for understanding human cognition.

Despite the theoretical debate, both approaches have outlined useful research methodologies for the field. In the next section, I will review some of the major methods of analyzing eye movement.

CHAPTER II

Methodology

The aim of this chapter is to survey multiple existing methodologies that describe, compare and model eye movement sequences. The chapter starts with a short review of the common methodologies used in the field and proceeds to the concept of scanpath—the ordered sequences of eye movement that consists of both temporal and spatial features of eye movement. Four current scanpath comparison methods are introduced and compared with the goal of finding a method that provides the best fit for the current eye movement data. Finally, a method that integrates multiple scanpath comparison approaches is proposed.

2.1 Issues of Traditional Eye Movement Methodology

Eye tracking is a dynamic methodology that produces large amounts of data that incorporate both spatial and temporal information. But the standard tools and methods in the field have rather limited capabilities that fail to take advantage of the wealth of information contained in eye movements (*Andrienko et al.*, 2012; *Mathôt et al.*, 2012).

Due to the complexity of eye movement data, the most common analysis methods often break the natural bond between spatial and temporal information and create a small set of separate unitary measures such as fixation duration (temporal) and areas of interest (spatial). As *Le Meur and Baccino* (2013) rightly put it, the field has primarily focused on synchronic indicators including fixation and saccade, rather than diachronic indicators such as scanpaths.

The synchronic approach treats an event as an occurrence at a specific point of time without taking history into account, while the diachronic approach considers the event as part of a process that develops over time (*Ramat et al.*, 2013). Because the movement sequences of eyes projected on to the environment they are looking can be seen as the trajectory of a moving journey, they can be represented by a continuous function of both time and space. Therefore it is reasonable to consider eye movements from a diachronic approach. By putting together both temporal and spatial information, the field can benefit from work in other research areas that provide more advanced methods capable of handling multiple dimensions simultaneously.

A second issue is that the common measures used in the field are highly aggregated across different spatial regions, time frames and individuals (*Coco*, 2009; *Ylitalo*, 2017). Often the abundance of information embedded in eye movement data is flattened to just a few single scores. The primary issues with such aggregated measures is that they oversimplify the variations between and within subjects. Because of the aggregation, such a measure will inevitably fail to capture the large variability and uncertainty that exist in any comparisons of eye movement. In sum, more comprehensive methods are needed to better represent the complexity of eye movement data.

2.2 Scanpath Comparison

Interest in eye movement sequences dates back to the early studies of *Noton and Stark* (1971b) and *Noton and Stark* (1971a), who recorded eye movements during pattern perception and found significant similarities between the sequence of fixations produced during initial inspection and subsequent presentation of the same visual stimuli. An ordered sequence of eye movement has been thereafter referred to as a *scanpath*, which is formally defined as the route of oculomotor events through space within a certain timespan (*Holmqvist et al.*, 2011c, p. 254).

Because eye movements are not single, independent events but rather reflect complex processes that unfold over time and space, we need measures that represent that complexity. Compared to synchronic indicators, scanpaths can better capture the spatial and temporal dimensions of eye movements during a task and are capable of revealing how visual processing plays out over time and space (*Noton and Stark*, 1971b,a). Unlike unitary eye movement events such as fixations and saccades, scanpaths incorporate multiple oculomotor events into one construct (*Dewhurst et al.*, 2012). Comparing different scanpaths can answer fundamental research questions such as whether an individual's eye movement has a relatively consistent pattern across different subjects and stimuli, whether an experimental manipulation is effective as reflected in systematic variation in eye movement sequences under different conditions, etc.

Raw scanpath data typically contain a set of fixations, the saccades connecting each of them, and the onset time and duration of these events. The data format has been similar across studies, but there exist various approaches to interpreting and representing scanpaths. For instance, scanpaths can be described as fixation maps, strings, geometric vectors or stochastic processes. Different representations have in turn motivated various methods for comparing scanpaths. They generally fall into four categories: 1) comparison of attention maps, 2) string edit methods, 3) geometric methods and 4) probabilistic approaches. The first three methods are more traditional and have mostly used similarity measure scores as the main way of comparing scanpaths. The motivation is to compare multiple scanpaths and return a single value that reflects the degree of similarity between eye movement sequences (*Mathôt et al.*, 2012). The last method, the probabilistic approach, is a more recent development. A cross-comparison of different scanpath comparison approaches can be found in Table 2.1. Next, I will discuss the strengths and weakness of each approach.

Table 2.1: Common Scanpath Comparison Methods

Method	Input	Output	Temporal Information	Predefined AOI
Attention map comparison	fixation location, fixation duration	correlation coefficient	No	Maybe
String edit	semantic label of fixation, fixation duration	similarity score	Yes	Yes
Geometric method	fixation location, fixation duration, saccade direction, saccade amplitude, etc.	similarity score	Yes	No
Probabilistic approach	fixation location, fixation duration, saccade direction, saccade amplitude, etc.	varies, most often classification model	Yes	No

2.2.1 Attention Map

An attention map is a heat map with hot spots representing either high fixation duration or frequently fixated areas. Other than the visualization form, an attention map also has mathematical representations that enable the calculation of correlation between different attention maps: stronger correlation indicates higher similarity. There are two common ways of generating an attention map from scanpaths: the AOI (Area of Interest) approach and the topographic approach.

The AOI approach segments the whole scene into a gridded area corresponding to areas of interest and then fills the grid with colors that represent total fixation durations or number of fixations of the particular cell (see Fig. 2.1 for example). There is no requirement that the areas of interest be of the same size.

Another approach to constructing an attention map is to use smoother, landscape-like representations with hills and valleys. This topographic approach requires creation of a one-to-one mapping of each point $S_i = (x_i, y_i)$ in the scanpath and the point on the attention map $G = f(x, y)$, where the most common mapping function $f(\cdot)$ is the Gaussian (see (2.1), σ_x and σ_y denote the horizontal and vertical standard deviation) (*Holmqvist et al.*, 2011c, p. 273), .

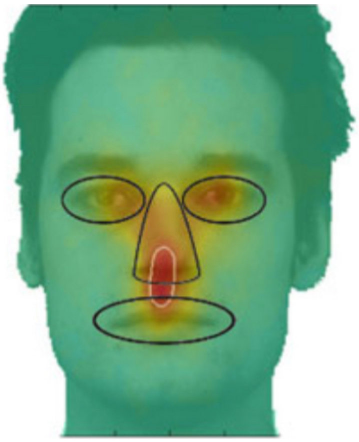

Figure 2.1: Example of Attention Map. Figure from *Caldara and Miellet* (2011).

$$f(x,y) = e^{-(\frac{(x-x_i)^2}{2\sigma_x{}^2}+\frac{(y-y_i)^2}{2\sigma_y{}^2})} \tag{2.1}$$

Different types of comparisons have been implemented based on these representations, including directly comparing individual fixation maps without considering the time dimension (*Caldara and Miellet*, 2011; *Leonards et al.*, 2007), or constructing a sequence of attention maps that partially retain order information (e.g., *Lao et al.*, 2015).

Attention maps produce visually attractive representations that have very high interpretability. This makes them a useful method for exploratory analyses as well as a widely used method for data-driven identification of AOIs. But an attention map is also an aggregated measure of similarity that is subject to oversimplification. A key shortcoming is that attention maps have been focused on the spatial dispersion of fixations at the expense of order or temporal information in the scanpath, making them sub-optimal for any research interested in the temporal properties of eye movement (*Mathôt et al.*, 2012; *Holmqvist et al.*, 2011b).

2.2.2 String Edit Method

The string edit method gets its name from 1) the recoding of raw eye movement sequences into AOI–based strings and 2) the similarity metric calculated from editing these strings. To be more specific, the entire scene is first divided into smaller areas (gridded AOIs or semantic AOIs); each of which is labeled with a character or group of characters. The eye movement sequence is thus transformed into an ordered list of characters. For example, for a classroom scene we could label students as S, instructional materials as I, a white board as B, the teacher as T etc. Then the original scanpath can be recoded as *SSBISTI*, with each letter denoting a fixation. Hence, these strings essentially represent an aggregated, simplified version of the actual scanpaths.

After recoding, the second step is to calculate the minimum number of character operations needed to match different strings. The unit cost of different character operations can be defined by various optimization algorithms. The most popular string edit algorithm in eye movement studies—Levenshtein distance—allows three single-character operations: insertion, deletion and substitution, each with a unit cost of 1 (*Levenshtein*, 1966; *Duchowski*, 2007a).

According to *Levenshtein* (1966), the distance between two strings is the minimum (optimum) number of single-character edits to transform one string into another. Arithmetically, Levenshtein distance is defined by Eq. (2.2).

$$ED_{R,S}(i,j) = \begin{cases} n & \text{if } m = 0 \\ m & \text{if } n = 0 \\ \min \begin{cases} ED_{R,S}(i-1,j)+1 \\ ED_{R,S}(i,j-1)+1 \\ ED_{R,S}(i-1,j-1)+I_A \end{cases} & \text{otherwise.} \end{cases} \tag{2.2}$$

$$I_a = \begin{cases} 1 & \text{if } R_i \neq S_j \\ 0 & \text{if } R_i = S_j \end{cases} \tag{2.3}$$

For example, consider two sequences *R: SSBISTI* and *S: SIBST*. In order to convert S into R, two matrices need to be constructed. The first is a substitution matrix that contains the edit cost between all the possible pairings of elements in R and S. According to Levenshtein distance, the edit costs are all 1s except for the diagonal where the row and column indicates identical letters (see Table 2.2). Then one can construct a comparison matrix with the two strings being compared forming the row and column divisions and each element being the unit cost of transformation (see Table 2.3). The comparison matrix uses the substitution matrix as a reference for finding the edit costs for each pair of letters. The emphasized value in the matrix is the minimal cost of partial transformation, while the cost to completely transform S to R can be found at the bottom right of the matrix. Thus the Levenshtein distance between R and S is 4.

Table 2.2: Substitution matrix for R and S according to Levenshtein distance

	B	I	S	T
B	0	1	1	1
I	1	0	1	1
S	1	1	0	1
T	1	1	1	0

Table 2.3: Comparison matrix for R and S. The optimal path is emphasized.

	S	I	B	S	T
S	0	1	2	3	4
S	1	*1*	2	2	3
B	2	2	*1*	2	3
I	3	2	2	*2*	3
S	4	3	3	3	3
T	5	4	4	4	*3*
I	6	5	5	5	*4*

The last step of the string edit method is to calculate the similarity measure. The similarity measure is often defined by the sum of costs along the optimal path, which is a simple value between 0 and 1 with 1 denoting complete match. For instance, the approach of *Privitera and Stark* (2000) first normalizes the total cost to the length of the longer string, yielding a sequence similarity index between two sequences of $S_s = (1 - 4/7) = 0.429$. The positional similarity index is calculated by comparing the characters from two strings; since all the characters in S are present in R, the above example yields a loci similarity index of $S_p = 1$.

Although Levenshtein distance is widely implemented in sequence comparisons due to its simplicity, there are some major issues that prevent its application in the current study. The most relevant one is that Levenshtein distance regards the distance between every character as equal, something not true in most cases. For example, comparing two string sequences: S1-S2 indicates when the teacher's fixation jumps from one student (S1) to another student (S2) who is physically adjacent to S1; and S1-B indicates when the teacher first fixates on a student (S1) then to the blackboard (B). It seems clear that the distance between S1 and S2 should be smaller than S1 and B, leading to a smaller edit cost. In addition to spatial proximity, the semantic content of the scene also introduces variations into the distance between different AOIs. For instance, saccades between students that are seated far apart and saccades between student and instructional material should have different distances. But the uniformally constructed Levenshtein distance substitution matrix will not reflect this difference; the editing cost of replacing S1 with S2, and replacing S1 with B is the same according to Levenshtein distance. Thus Levenshtein distance is not the optimal distance algorithm when the AOIs are not uniformally distributed in the scene. Also, given that eye movements are essentially time series data, the Levenshtein distance approach has converted the time sequence to ordinal data by discarding all the fixation duration information.

To solve these two issues, *Cristino et al.* (2010)'s ScanMatch provides a more advanced adaptation of Levenshtein distance using the Needleman-Wunsch algorithm and temporal binning. Another practical improvement of ScanMatch is to implement double-string coding instead of single string to increase the number of possible AOIs.

The first change ScanMatch implements is to incorporate fixation time by using temporal binning—repeating the letter corresponding to an AOI in a way that is proportional to the fixation duration within that AOI. For instance, the sequence STB might turn into SSSSTTBBB with 50 millisecond bins. In this way, the coded string will incorporate spatial location, sequential information, and temporal duration of fixations.

The second change is that ScanMatch takes into account the relationship between AOIs and specifies the alignment score accordingly by using a sequence alignment algorithm borrowed from the field of bioinformatics: the Needleman-Wunsch algorithm. It does so in the following way. Similar to the classic string edit approach, a substitution matrix needs to be constructed. But instead of weighting every change equally, the score for editing between two letters can be defined according to some measure of the relationship between AOIs. This can be weighted so that a higher score indicates better alignment. The alignment score can be defined according to the Euclidean distance between bins, physical properties (color, size, shape, etc.) or semantic relationships (human, non-human objects, child, adult, etc.). Another parameter that determines scoring is called the gap penalty. The value of the gap penalty indicates the score for alignment of a letter with a gap instead of another letter. It essentially serves as a threshold for making the choice of either performing alignment or inserting a gap that hurts local alignment but may benefit the global alignment. Different gap penalty values will influence the behavior of algorithm in finding the optimum route, for example, if we have two sequences sAsB and sAsC with substitution matrix 2.4. If the gap penalty equals 0, the score of aligning sA and sB is higher than the gap penalty and therefore a gap is inserted. If the gap penalty is higher than the cost of aligning two letters, then string alignment is favored over insertion of gap (see Fig. II.1). *Cristino et al.* (2010) argued that if the alignment score in the substitution matrix is well chosen then the gap penalty can be set to zero. Providing the information of substitution matrix and gap penalty, the optimal route with highest alignment score is sought within the comparison matrix. The final score is normalized over the product of the max alignment score and the length of the longer sequence and gives a single similarity measure between 0 and 1, with 1 representing perfect match.

Table 2.4: Substitution Matrix

	sA	sB	sC
sA	10	3	-2
sB	3	10	-5
sC	-2	-5	10

```
If gap penalty = 0: sA>>sA = 10, gap>>sC = 0, sB>>gap = 0
    then overall score = 10 + 0 + 0 = 10
If gap penalty = -6: sA>>sA = 10, sB>>sC = -5>-6
    then overall score = 10 + (-5) = 5
```

Listing II.1: Compare String Alignment and Gap Insertion

Although it is the best string edit method so far, ScanMatch still suffers from the limitation of only producing aggregated results. The single similarity score prevents further exploration and interpretation of the result: it provides answers to “how different these scanpaths are” but not “where do the differences reside” or, moving one step further, “how important is this difference to the overall results”. These questions are better addressed with other scanpath comparison methods.

2.2.3 Geometric Method

In contrast to the string edit method's approach of defining a priori AOIs and recoding spatial-temporal data as ordered strings, geometric methods avoid recoding by representing eye movements according to their geometric properties such as location coordinates, fixation duration, saccade direction, saccade velocity, saccade amplitude, etc.

Compared to AOI-based methods, geometric representation has the merits of retaining the vector properties of eye movements, such as saccade direction and scanpath shape, despite scaling and shifting. Also, geometric methods use raw, continuous data instead of partitioned, coarser data, thus avoiding the quantization error AOI segmentation introduces to the analysis. These properties make them superior to AOI based methods when geometric properties are of interest.

Mannan et al. (1995) were among the first to represent eye movements by location coordinates without defining AOIs. They proposed a linear distance method that calculates the Euclidean distances between every pair of nearest neighbours in two scanpaths (*Mannan et al.*, 1995, 1996; *Mathôt et al.*, 2012; *Dewhurst et al.*, 2012). Firstly, each fixation in one set is mapped onto another fixation from the other set that is closest in terms of location coordinates. Then a point-mapping distance (Euclidean distance) $d_M(\mathbf{r}, \mathbf{s})$ between a point (fixation location) $\mathbf{r}$ in an eye movement sequence R and its nearest neighbour $\mathbf{s}$ in another eye movement sequence S will be calculated (see Eq. (2.4)). In this case, $\mathbf{r}$ and $\mathbf{s}$ are two points in a two-dimensional Euclidean space with coordinates (r_1, r_2) and (s_1, s_2).

$$d_M(\mathbf{r}, \mathbf{s}) = \sqrt{(r_1 - s_1)^2 + (r_2 - s_2)^2} \tag{2.4}$$

The mapping and point-mapping distance calculation is repeated in the other direction as well, namely mapping each $\mathbf{s}$ from S onto the nearest neighbor $\mathbf{s}$ from R.

This technique is called *double-mapping*; *Mathôt et al.* (2012) argues that double-mapping is computationally cheap and at the same time not inferior to more sophisticated heuristics. Also, by performing the same mapping process twice, double-mapping is capable of comparing sequences that are discrepant in length. The final similarity measure of this process is called sequence-mapping distance; this is the collection of all the point-mapping distances normalized by the total length (number of points) of both sequences, this sequence-mapping distance $D_M(R,S)$ is defined by Eq. (2.5). Here, n_R and n_S is the length of R and S respectively, $d_M(r_i,)$ is the distance between the i^{th} point in R and its nearest neighbour in S, likewise $d_M(s_j,)$ is the distance between the j^{th} point in S and its nearest neighbor in R. The work-flow of this method can be summarized in pseudo-code form (see Listing II.2).

$$D_M(R,S) = \frac{\sum_{i=1}^{n_R} d_M(r_i,)^2 + \sum_{j=1}^{n_S} d_M(s_j,)^2}{sum(n_R,n_S)} \tag{2.5}$$

```
BEGIN
D = 0
FOR all points r in sequence R:
    FIND nearest neighbour s in sequence S
    D = D + Euclidean distance(r,s)
FOR all points s in sequence S:
    FIND nearest neighbour r in sequence R
    D = D + Euclidean distance(r,s)
D = D / sum(size(R), size(S))
END;
```

Listing II.2: Pseudocode for Mannan's algorithm

As Eq. (2.4) and Eq. (2.5) show, the Mannan linear distance approach can only handle two-dimensional coordinates and doesn't take any other eye movement information into account (such as fixation duration, fixation order etc.). To address this concern, *Mathôt et al.* (2012) extends Mannan's method to multiple dimensions. Their Eyenalysis method employs multidimensional Euclidean distance and the fixation can be defined by any number and combination of dimensions. For example, instead of only including location coordinates, fixations can now be defined in terms of onset time and duration as well: (x, y, t, d). The algorithm is similar to *Mannan et al.* (1995)'s method but incorporates multiple dimensions. The point-mapping distance $d_E(\mathbf{r}, \mathbf{s})$ is defined by Eq. (2.6), with $\mathbf{r}$ and $\mathbf{s}$ representing a point in sequence R and S that resides in an n-dimensional Euclidean space. In this representation, n denotes the number of dimensions while s_i and r_i indicate the value on the i^{th} dimension of points $\mathbf{r}$ and $\mathbf{s}$ respectively. Following the same method as *Mannan et al.* (1995), the final similarity measure—sequence-mapping distance is defined by (2.7), with a slight difference of normalizing the sum of point-mapping distances by the number of points in the longer sequence instead of normalizing by the total number of points from both sequences. Likewise, the Eyenalysis method can also be intuitively expressed in a pseudo-code form (see Listing II.3).

The differences between Mannan's and Eyenalysis's approach lies in a) Number of dimensions that were taken into account. This can be seen by comparing Eq. (2.4) and Eq. (2.6). In Eq. (2.4), $\mathbf{r}$, $\mathbf{s}$ are two points in a two-dimensional Euclidean space, while in Eq. (2.6) $\mathbf{r}$, $\mathbf{s}$ represents a point in sequence R and S that both resides in an n-dimensional space. b) Ways of normalization. Mannan's distance is normalized over the total length and Eyenalysis is normalized over the longest length among two sequences.

$$d_E(\mathbf{r}, \mathbf{s}) = \sqrt{\sum_{i=1}^{n} (r_i - s_i)^2} \tag{2.6}$$

$$D_E(R,S) = \frac{\sum_{i=1}^{n_R} d_E(r_i,) + \sum_{j=1}^{n_S} d_E(s_j,)}{max(n_R, n_S)} \tag{2.7}$$

```
BEGIN
D = 0
FOR all points r in sequence R:
    FIND nearest neighbour s in sequence S
    D = D + Euclidean distance(r,s)
FOR all points s in sequence S:
    FIND nearest neighbour r in sequence R
    D = D + Euclidean distance(r,s)
D = D / max(size(R), size(S))
END;
```

Listing II.3: Pseudocode for Eyenalysis's algorithm, adapted from Mathot et al., (2012)

The basis of these two methods is finding the nearest neighbor based on location coordinates, thus yielding reliable performance in quantifying location similarities/dissimilarities. But on the other hand, they lack sensitivity in identifying other geometric properties such as scanpath shape and direction.

Dewhurst et al. (2012) proposed a more flexible geometric method called MultiMatch. It is designed to compare multiple scanpath dimensions including order, position, shape, saccade length, direction and fixation duration by representing scanpaths in a vector form (*Jarodzka et al.*, 2010b,a; *Dewhurst et al.*, 2012). *Dewhurst et al.* (2012) argued that "ideal saccades" (saccades that take the shortest route between two fixations) can be expressed as a Euclidean vector $\mathbf{r} = x, y$, a mathematical entity with magnitude x and direction y. The scanpath being compared can be denoted as $R = \mathbf{r_1}, \mathbf{r_2}, \cdots, \mathbf{r_i}, \cdots, \mathbf{r_m}$ and $S = \mathbf{s_1}, \mathbf{s_2}, \cdots, \mathbf{s_j}, \cdots, \mathbf{s_n}$.

MultiMatch is conducted in three steps (see Table 2.5 for summary). First, unimportant local eye movements are merged and replaced with a new vector defined as the sum of these local vectors. More specifically, both the amplitudes and directions of consecutive vectors are compared against arbitrary thresholds T_{amp} and T_{angle} (usually defined as 10 percent of the screen diagonal and 45 respectively). If the group of vectors is smaller than the threshold along these two dimensions they will then be replaced with a new grouping vector $\mathbf{r} = \mathbf{r_1} + \mathbf{r_2} + \; + \mathbf{r_m}$.

Table 2.5: MultiMatch Procedure

Step	Data Dimensions Used	Method Used
Simplification	saccade amplitude and saccade direction	amplitude-based clustering and direction-based clustering
Alignment	usually shape	finding the shortest path based on graph theory
Comparison	shape, fixation location, fixation duration, saccade amplitude (length), saccade direction	significance test

Similar to the string edit approach, MultiMatch also requires aligning two sequences first before comparing their differences. As with string editing, a comparison matrix that represents the cost accompanying the pairwise comparison of all the elements in two sequences needs to be constructed. While the string edit approach defines the cost by Levenshtein distance, MultiMatch uses vector differences instead. Specifically, the difference between two vectors $\mathbf{r_i}$ and $\mathbf{s_j}$ from two vector sequences R and S is represented by the magnitude of the differential vector $\mathbf{r_i} - \mathbf{s_j}$, i.e., $||\mathbf{r_i} - \mathbf{s_j}||$. Table 2.6 is the example comparison matrix used by *Dewhurst et al.* (2012), with each cell containing the length of the differential vector between any two saccadic vectors, expressed in degrees of visual angle. The cost gets smaller when two vectors are similar in magnitude and direction, signaling a good alignment between segments of two scanpaths.

In order to implement Dijkstra's algorithm (*Dijkstra*, 1959), a comparison matrix is then used to construct an adjacency matrix and subsequently a directed *graph*. The *graph* is a mathematical abstraction of situations, consisting of a set of points

Table 2.6: MultiMatch Comparison Matrix

	$\mathbf{r_1}$	$\mathbf{r_2}$	$\mathbf{r_3}$	$\mathbf{r_4}$	$\mathbf{r_5}$
$\mathbf{s_1}$	8	9	6	7	5
$\mathbf{s_2}$	8	1	10	12	11
$\mathbf{s_3}$	4	12	0.5	7	3
$\mathbf{s_4}$	7	12	7	1	4
$\mathbf{s_5}$	4	11	1	7	4

together with lines joining certain pairs of them (*Bondy et al.*, 1976, p. 1). Graphs are used to model pairwise relations between objects; the point is called a *vertex* and the connecting line an *edge* in this context. These vertices can represent people in a social network, locations in navigation systems, steps in project management, or in our case, elements of a comparison matrix etc. The graph can be directed or undirected depending on whether the edge has a specific direction. Because the question of aligning two scanpaths is essentially one of finding an optimum path, with a start and end point, through a comparison matrix, the comparison matrix should be modeled as a directed graph, or digraph. Also, each edge is associated with a weight in a graph, also referred to as cost or distance, defined as the vector differences between two saccadic vectors in the current case (see Table 2.7 for a summary of terms).

Table 2.7: Summary of Graph Theory Terms

Graph Theory Term	Definition	Application in MultiMatch
Vertex	A finite set of points	Elements in a comparison matrix
Edge	Links connecting pairs of vertices	Adjacency indicators
Weight	Numeric values assigned to edges	Vector differences (can be other similarity matrices as well)
Shortest path problem	Finding a path between two vertices that minimizes the sum of weights	Finding the best alignment between scanpaths
Path	An alternating sequence of vertices and edges	Directional calculations of the optimum cost

In preparation for transforming the comparison matrix into a graph, the allowed transitions among comparison matrix elements need to be specified. They will serve as the links between matrix elements. Under the requirement that this matrix can be recreated as a digraph with the first element as the start vertex and last element the end vertex, the represented direction of path thus travels from top-left to bottom-right with no retracing of steps. Then all the possible transitions during this travel can be identified. Following the travel metaphor, allowed transitions can be thought as neighborship (adjacency) indicators, illustrating whether two vertices are connected along the directed path. Traditionally, an adjacency matrix is a square matrix that summarizes the relationship between vertices, with elements indicating the number of edges connecting two vertices. For instance, 0 means the pair of vertices have no connecting edges or are not adjacent. MultiMatch has used a weighted adjacency matrix that stores edge weights directly in the elements to demonstrate the rules for how comparison matrix elements are connected; the weight is the same as the similarity metric (vector difference) used in the comparison matrix. For example, a comparison matrix (2.8) and the corresponding adjacency matrix (2.9) are presented here:

$$M_{m,n} = \begin{array}{cc} & \begin{array}{ccccc} \mathbf{s_1} & \cdots & \mathbf{s_j} & \cdots & \mathbf{s_n} \end{array} \\ \begin{array}{c} \mathbf{r_1} \\ \vdots \\ \mathbf{r_i} \\ \vdots \\ \mathbf{r_m} \end{array} & \begin{pmatrix} w_1 & \cdots & w_j & \cdots & w_n \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{(i-1)m+1} & \cdots & w_{(i-1)m+j} & \cdots & w_{(i-1)m+n} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ w_{(m-1)m+1} & \cdots & w_{(m-1)m+j} & \cdots & w_{(m-1)m+n} \end{pmatrix} \end{array} \tag{2.8}$$

$$A_{mn,mn} = \begin{array}{cc} & \begin{array}{ccccc} 1 & \cdots & l & \cdots & mn \end{array} \\ \begin{array}{c} 1 \\ \vdots \\ k \\ \vdots \\ mn \end{array} & \begin{pmatrix} 0 & \cdots & w_{1,l} & \cdots & w_{1,mn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & w_{k,l} & \cdots & w_{k,mn} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & w_{mn,l} & \cdots & w_{mn,mn} \end{pmatrix} \end{array} \tag{2.9}$$

When the k^{th} element in one scanpath and l^{th} element in another scanpath are adjacent, then the value of adjacency matrix $w_{k,l}$ equals the l^{th} element in the comparison matrix, other wise is 0. The adjacency relationship between any two elements in the comparison matrix can be found using the pseudo-code II.4.

```
BEGIN
FOR each row $i(i \in 1, m)$ in the comparison matrix $M_{m,n}$
    FOR each column $j(j \in 1, n)$ in the comparison matrix $M_{m,n}$
        IF $j \leq i$ THEN
            $w_{(i-1)m+j} = 0$
        ELSE
            IF $i \neq m$ THEN
                IF $j \neq n$ THEN
                    $w_{(i-1)m+j}$ is adjacent to three elements: $w_{(i-1)m+j+1}$,
                        $w_{(i-1)m+m+j}$, $w_{(i-1)m+n+j+1}$
                ELSE
                    $w_{(i-1)m+n}$ is adjacent to one element: $w_{(i-1)m+2n}$
                END IF
            ELSE
                IF $j \neq n$ THEN
                    $w_{(m-1)m+j}$ is adjacent to one element: $w_{(m-1)m+j+1}$
                ELSE
                    $w_{(m-1)m+n} = 0$
                END IF
            END IF
        END IF
    END FOR
END FOR
```

```
END;
```

Listing II.4: Pseudocode for finding adjacency relationship

A graph representation of the shortest path problem (aligning two scanpaths) can then be created based on the adjacency matrix. MultiMatch then applies Dijkstra's algorithm to find the optimal path that minimizes the cost, or in graph theory terms, finds the shortest path between the first and last vertex. This path defines how the vectors in two scanpaths can be aligned, namely ensuring that each vector in one scanpath is matched with a vector in the other scanpath. For example if the shortest path is vertex 1-4-5, indicating $\mathbf{r_1}$ is matched with $\mathbf{s_1}$, $\mathbf{r_2}$ with $\mathbf{s_2}$, and then $\mathbf{s_3}$ with $\mathbf{r_2}$.

Finally similarity is calculated on the aligned scanpaths with respect to five different aspects of the scanpaths: shape, amplitude, direction, position and duration (see Table 2.8). Each category yields an averaged value over all the matching pairs. This value is normalized with its largest possible value and then inverted to obtain an interval of $0, 1$ with 1 represents perfect match and 0 represents no similarity.

Table 2.8: MultiMatch Similarity Measure

Measurement	Normalization	Method
Shape	Screen diagonal	Compute vector difference between the aligned saccade pairs $\mathbf{u_i} - \mathbf{v_j}$
Amplitude (length)	Screen diagonal	Compute distance between the endpoints of saccade vectors
Direction	π	Compute angular difference between saccade vectors
Position	Screen diagonal	Compute Euclidean distance between aligned fixations
Duration	The maximum duration among the two	Compute difference in fixation durations between aligned fixations

MultiMatch succeeds in preserving all the geometric properties of scanpath, and provides more elaborated information on the type of (dis)similarities given a particular dimensionality (*Le Meur and Baccino*, 2013). It is superior to the string edit method in the sense that possible dimensions extend beyond Euclidean distance between fixations to include overall shape, saccade direction, saccade length, and fixation duration. But it inevitably falls short of ScanMatch in that the alignment procedure is blind to the semantic meaning of the fixations. The raw location coordinates are not labeled and thus it may be difficult to interpret the (dis)similarities.

2.2.4 Probabilistic Approach

The essential characteristic of probabilistic approaches is the assumption that eye movement events are random variables manifested as the observable outcomes of underlying stochastic processes (*Boccignone*, 2017; *Coutrot et al.*, 2017). With this assumption a wider variety of advanced methods become available, especially techniques borrowed from the field of statistical machine learning.

A stochastic process can be formally defined as a collection $\{X_t; t \in T\}$ of random variables X_t (also, $X(t)$) defined in the same probability space with and taking values from *state space* S. X_t are indexed by *parameter set* T which customarily represents time (either discrete or continuous). Then X_t can be thought as the *state* of the process at time t (*Cinlar*, 2013; *Boccignone*, 2017). Applying this definition to eye movements we get $X_{t_i} = x_i$ with x_i being the observed eye movements and realizations of the stochastic process. x_i may contain multiple dimensions such as fixation location, saccade amplitude and direction. Then the mapping from visual input I to a sequence of eye movements under a certain task $\mathbf{T}$ can be simply written as $I \underset{\mathbf{T}}{\rightarrow} \{x_{(1)}, x_{(2)}, x_{(3)}, \cdots\}$. Here I represents the features of raw visual input such as colors and shapes, and $\mathbf{T}$ can be any tasks that requires cognitive effort such as visual exploration, signal search, face recognition, navigation, etc.

2.2.4.1 Bayesian vs. Frequentist Framework

A large body of human eye movement studies have been concerned with the question, *where do people look?* A wide array of computational models such as the feature integration theory of *Treisman and Gelade* (1980), the selective routing model of *Koch* *and Ullman* (1987), the shifter circuit model of *Olshausen et al.* (1993), the temporal tagging model of *Niebur et al.* (1993), and the selective tuning model of *Tsotsos et al.* (1995), just to name a few, have been designed to untangle this question with no completely satisfactory results (*Heinke and Humphreys*, 2005). Most of these models were built upon the saliency representations of visual inputs, namely the low-level visual features (*Koch and Ullman*, 1985; *Itti et al.*, 1998), while seldom incorporating the influence of task.

Up to now, no models have really succeeded in predicting natural eye movement sequences when looking at arbitrary scenes (*Frintrop et al.*, 2010; *Boccignone*, 2017). Using Receiver Operating Characteristic analysis, *Tatler and Vincent* (2009a) demonstrated that models with terms learned from actual eye movement outperform the traditional saliency models built solely upon features of visual input: .648 as opposed to .565 (area under the receiver operator curve; larger area indicates better classification performance). The natural follow-up question is where does the difference in model performance come from? Aside from the apparent complication of modeling a delicate process that is shaped by both voluntary and involuntary forces, a major limitation of the current modeling approaches can be exposed by simply examining their mathematical specifications.

Tatler and Vincent (2009a) showed that the fundamental research question—*where do people look?*—provided certain visual input can be rephrased as the problem of finding the conditional probability of an eye movement sequence given information about visual input: $P(x \mid I)$. Given the definition of conditional probability (2.10) and the Multiplication Law (2.11), the probability $P(x \mid I)$ can be written as (2.12), which is essentially Bayes' theorem (*Tatler and Vincent*, 2009a; *Boccignone*, 2017). Thus we know the question can be interpreted under the Bayesian framework, as the prior distribution of eye movements $P(x)$ being updated by the sampled data about visual features at a given fixation location $P(I \mid x)$ normalized by the overall likelihood of these features occurring in the environment $P(I)$ (*Boccignone*, 2017).

$$P(E \mid F) = \frac{P(E \cap F)}{P(F)} \tag{2.10}$$

$$P(E \cap F) = P(E \mid F)P(F) = P(F \mid E)P(E) \tag{2.11}$$

$$\overbrace{P(x \mid I)}^{\text{posterior prob. of saccades}} = \frac{P(x \cap I)}{P(I)} = \frac{\overbrace{P(I \mid x)}^{\text{prob. of feature given certain saccades}}}{\underbrace{P(I)}_{\text{prob. of features occurring in the environment}}} \overbrace{P(x)}^{\text{prior prob. of saccades}} \tag{2.12}$$

Now that we have the Bayesian representation of the research question *where do people look*, let's go back to the traditional visual attention model—the saliency model. *Boccignone* (2017) has pointed out, despite the abundance of versions of these models, most saliency models were built on a core representation Eq. (2.13), a form that is similar with Eq. (2.12), but discarded information about transitions between consecutive fixations $x_f(t) \to x_f(t+1)$, essentially modeling on prior probability of fixations instead of saccades (e.g., *Borji and Itti*, 2013). Most often, Eq. (2.13) is further simplified as Eq. (2.14) by setting $P(I \mid x_f)$ and $P(x_f)$ to constant. Eq. (2.14) can be interpreted as: the probability of fixation at certain location is proportional to the salience of visual features at that location; smaller $P(I)$ indicates the feature is less likely to appear in the environment, and is thus more salient due to its being unexpected. *Tatler and Vincent* (2009a) criticized the choice of dropping $P(x_f)$ in (2.14), arguing that it shows a neglect of inherent biases/systematic tendencies of real eye movements irrespective of the visual input (*Tatler et al.*, 2011; *Tatler and Vincent*, 2009a). They argued that human eye movements evince natural systematic tendencies that are independent of the external environment including conspicuities of visual stimulus. The tendencies they discovered include more horizontal saccades than vertical ones, a larger probability of fixating on central areas, among others. *Tatler and Vincent* (2009a)'s analysis provides an alternative approach to modeling eye movements that should be explored in the future, starting with stationary eye tracking paradigms.

$$\overbrace{P(x_f \mid I)}^{\text{posterior prob. of fixations}} = \frac{\overbrace{P(I \mid x_f)}^{\text{prob. of feature given certain fixations}}}{\underbrace{P(I)}_{\text{prob. of features occurring in the environment}}} \overbrace{P(x_f)}^{\text{prior prob. of fixations}} \tag{2.13}$$

$$\overbrace{P(x_f \mid I)}^{\text{posterior prob. of fixations}} \propto \frac{\overbrace{1}^{\text{salience at given locations}}}{\underbrace{P(I)}_{\text{prob. of features occurring in the environment}}} \tag{2.14}$$

By comparing Eq. (2.12) and Eq. (2.14) we can see that they are derived from two different interpretations of the distribution of eye movements. Eq. (2.14) treats fixation location as a fixed value rather than a distribution that has dependency on past events. This corresponds to the frequentist view of observed data, while Eq. (2.12) follows the Bayesian framework and considers eye movements as a dynamic process.

Based on this comparison of two frameworks, the current paper will adopt the Bayesian interpretation of probability given the dynamic nature of eye movements. Specifically, the current paper considers the eye movement sequence $\{x(t=1), x(t=2), \cdots\}$ as a sample from a stochastic process with probability density function $P(x)$.

2.2.4.2 Examples of Probabilistic Approach: Hidden Markov Models

As discussed before, the dynamic nature of eye movement is of central focus in the current paper. To more formally define the *dynamics* of the stochastic process in order to fully describe the statistical properties of the scanpath we have the *tran-*

sition probabilities representing the probability of state transitions. Thus an n-step transition probability P_{ij}^{n} is the probability that a process in state i will be in state j after n transitions as in Eq. (2.15) (*Ross*, 2014, p. 187). Thus the probability of the process at time $k+l$ can be written as a joint probability (2.16) given information from all past time points and the current time.

$$P_{ij}^{n} = P\{X_{n+k} = j \mid X_k = i\}, n \geq 0, i, j \geq 0 \tag{2.15}$$

$$P\left(x_1, t_1; \cdots; x_{k+l}, t_{k+l}\right) = P(x_{k+1}, t_{k+1}; \cdots; x_{k+l}, t_{k+l} \mid x_1, t_1; \cdots; x_k, t_k) \cdot P(x_1, t_1; \cdots; x_k, t_k) \tag{2.16}$$

A Purely Random Process is the simplest stochastic process in the sense that $X(t)$ is completely independent of past or future values; thus the joint probability $P\left(x_1, t_1; \cdots; x_{k+l}, t_{k+l}\right)$ is simply the product of all $P(x_i, t_i)$. But in reality, a stochastic process is often not as simple as this. Going one step further to describe a more realistic situation yields a *Markov process* in which each state is memoryless beyond the most recent transition, namely Eq. (2.17)

$$P\left(x_n, t_n \mid x_{n-1}, t_{n-1}; \cdots; x_1, t_1\right) = P(x_n, t_n \mid x_{n-1}, t_{n-1}) \tag{2.17}$$

The Markov process is the most widely-used probabilistic description of eye movement sequences (e.g., *Liechty et al.*, 2003; *Feng*, 2006; *Simola et al.*, 2008; *Kit and Sullivan*, 2016; *Coutrot et al.*, 2017), and Hidden Markov models are representatives of this approach. A Hidden Markov model (HMM) is a stochastic model in which the process (time series data) being modeled is assumed to be a Markov process with discrete, hidden states. HMM is designed to represent how qualitatively different, unobserved states unfold over time (*Visser*, 2011; *Baum and Petrie*, 1966).

Before explaining HMM, some notations need to be clarified. Consider observations collected over time length t and denote the data observed at time t by the variable Y_t; thus the time series data are denoted by $Y_{1:T} := (Y_1, Y_2, \cdots, Y_T)$. Also consider the state variable S_t that represents the hidden process generating the observations, $S_{1:T} := (S_1, S_2, \cdots, S_T)$.

HMM has three defining characteristics:

a) States are discrete: the marginal distribution of the time series data is a mixture distribution. In other words, the data do not follow any single uni-modal distribution but instead are sampled from multiple distributions with different parameters. In a more rigorous definition, HMM is said to be generated by multiple discrete states or components; this can be represented as Eq. (2.18).

$$f(Y_t) = \sum_{i=1}^{n} p_i f_i(Y_t) \tag{2.18}$$

Here p_i are the component proportions, and the marginal density function of data is the sum of all the conditional distribution $f_i(\cdot)$ of the data. Also, since the states are discrete, S_t can take on values from a set of integers denoted by $K = \{1, \cdots, n\}$. K is called the state-space of HMM and n is the number of discrete states.

b) States are hidden: HMM is different from observable Markov models in that the states generating observations are hidden from the observer. That is to say, the mapping between S_t and Y_t is probabilistic rather than deterministic.

c) States follow a Markov process: HMM assumes that the state variable S_t satisfies the Markov property, that is, stating that the conditional distribution of any current state S_t, given the previous state S_{t-1}, is independent of the all past states prior to $t-1$ and depends only on the previous state S_{t-1}, expressed as (2.19).

$$P\left(S_t \mid S_1, S_2, \cdots, S_{t-1}\right) = P\left(S_t \mid S_{t-1}\right) \tag{2.19}$$

In sum, HMM is characterized as a model with discrete, hidden states, each with a different distribution function, and the evolution of states over time follows a Markov process (*Ghahramani*, 2001; *Visser*, 2011).

HMM has been used in psychology to represent situations where the states described above are assumed to be mental states that characterize typical human behavior. This list includes sleep stages, problem solving strategies, grammatical rules, covert visual attention modes, the kind of task one is doing, an observer's expertise, emotional state, etc. (*Flexer et al.*, 2000, 2005; *Schmittmann et al.*, 2005; *Visser et al.*, 2002; *Liechty et al.*, 2003).

The application of HMM in eye movement research is still at an early stage. The few existing studies have presumed that scanpaths are better described by a stochastic process rather than a deterministic sequence, essentially proposing that scanpaths exhibit a Markov process in which the position of a fixation depends (and only depends) on the position of the previous fixation (*Viviani*, 1990; *Ellis and Smith*, 1985; *Liechty et al.*, 2003).

In related eye movement research, there are two approaches using HMMs. The first approach clearly defines the hidden states as modes of visual attention. For instance, *Liechty et al.* (2003) postulated that people switch between two distinct covert visual attention states, namely local and global visual attention during visual scanning. Local visual attention enables better extraction of details from specific and adjacent locations. Global visual attention facilitates integration of information from multiple locations and is used for identifying next locations to fixate. *Liechty et al.* (2003) thus presumed that local and global attention are associated with different length of saccades: local attention is characterized by shorter saccades while longer saccades represent global attention. This is not the most suitable approach for the current data set for two reasons. First, the presumed association between visual attention states and saccade length has little empirical support. More importantly, making inferences about attention states is not the primary goal of the current paper.

Another approach regards HMM as a *generative model* and has only begun to appear in recent years. This approach has explicitly or implicitly modeled the distribution of eye movement (input) around the hidden state as well as the transition matrix of hidden states (output) and it is also capable of generating synthetic data in the same input space, hence the name *generative model* (*Boccignone*, 2017; *Coutrot et al.*, 2017). For this approach, an area of interest (AOI) is not a collection of fixations that falls in that area (or in other words, fixations were not equated to the exact points of interest), but rather a distribution sampled from some states that are not directly observable. And HMM is used to model both the distribution of actual observations and the hidden states. For example, *Coutrot et al.* (2017) and *Chuk et al.* (2014) have assumed a hidden state to be the region of interest (ROI) in the image and observed data to be fixation locations. They modeled the emission densities (the distribution of fixations in each ROI) as a 2D Gaussian distribution, that is, given a hidden state the actual observation is assumed to be sampled from a Gaussian distribution centered around the true *point of interest*. The transition matrix of hidden states can be inferred from the input data and the parameters defining HMM is chosen based on the magnitude of likelihood for a new input observation. I will further discuss this approach under the frame work of Bayes' Theorem in a later section.

2.3 Challenges of the Current Data

The current eye movement data are complicated due to factors inherent in the nature of teaching. Teaching is a cognitively complex, goal-oriented task involving simultaneous processes that occur in a complex environment with multiple interacting actors and a variety of instruction-related stimuli, not to mention distractions. As already mentioned, goal-oriented and free-viewing tasks are used to examine the relative impact of bottom-up and top-down processes, but they are rarely combined in non-lab settings. The variables that characterize a complex system, such as real-life natural eye movements, are difficult to control and isolate (*Keane et al.*, 2014). Thus the complexity of natural eye movements requires more advanced statistical methods.

The largest challenge is probably how to represent quantitatively where the teacher is looking. Since the teachers who participated in the current study were wearing a mobile eye tracker that does not track head position or physical movement with respect to the external environment, we have no way of knowing the exact location coordinates as we would in traditional static scene perception studies. As we have reviewed in past sections, most scanpath comparison methods utilize the two-dimensional coordinates on screen to define gaze location, but the lack of a consistent coordinate system precludes a straightforward application of this method. The next section will describe how we can overcome this obstacle by transforming the original eye movement data.

2.4 General Analysis Framework

Our core research question involves what we can learn from teacher's eye movements that will tell us something about the cognitive processes involved in teaching as they relate to teacher attention. Our goal is to analyze scanpath patterns and further make inferences about the scanpath based on eye movement sequences. We collected data comparing novice and experienced teachers teaching the same students and elementary school teachers teaching different subjects (reading and mathematics) to the same students. We are interested in how expertise and subject matter affect patterns of looking.

Both of these questions can be rephrased as an inverse problem under Bayes' framework: based on eye movement data, can we identify whether the participant is a novice or experienced teacher? Similarly, can we identify whether a given teacher is teaching mathematics or literacy?

How to solve the inverse problem is the key step in analyzing our data. Next, I will examine my research questions under the framework of statistical learning.

2.4.1 Problem of Classification

As we covered in the previous section 2.2.4, the basic assumption of the probabilistic approach is that some hidden mental states Y have shaped the eye movement X we observe, represented by the generative process $Y \to X$. In practice, we can not observe this process directly but we can make inferences about the target states based on the relationship between Y and X, namely, finding the probability of the target state given the observed eye movement data. This inference problem $X \to Y$ can be written as Eq. (2.20) by applying Bayes' Theorem and the multiplication law of conditional probability $P(Y, X) = P(X \mid Y)P(Y)$. $P(Y, X)$ is the posterior probability we can use to determine the output $Y = y$ for each new input $X = x$. The parameters specifying the probability density functions in Eq. (2.20) can be learned from the actual observations, and to this end many statistical learning techniques can be utilized.

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \tag{2.20}$$

This inferential process is of central focus in the area of statistical learning, which deals with the problem of finding the relationship captured by predictive functions between input data and output variables (*James et al.*, 2013; *Friedman et al.*, 2001). Depending on the relationship between input and output, as well as the data type of output, statistical learning problems can be roughly categorized into four kinds (Table 2.9). In *supervised learning*, observations x_i are associated with target values y_i (also called labels, response measurements, or dependent variables), or in other words, the input data is labeled with target values. On the other hand, *unsupervised* *learning* deals with unlabeled input data with no "supervision" from the target values.

Table 2.9: Types of Statistical Learning Problems

	Supervised Learning (Labeled Data)	Unsupervised Learning (Unlabeled Data)
Y is Discrete	Classification	Clustering
Y is Continuous	Regression	Dimensionality Reduction

The goal of supervised learning is to understand the relationship between X and Y in order to predict the target value of a new input. Depending on the data type of output, supervised learning can be further broken down to a *classification problem* or *regression problem*. When the output values belong to a discrete set of labels, it is a classification problem; when the output takes a continuous range of values, the problem is known as regression (*Boccignone*, 2017; *Friedman et al.*, 2001).

With no labels attached, the main goal of unsupervised learning then becomes mining for interesting patterns that are yet unknown. Similarly, when Y is discrete the problem is called *clustering* and if Y is continuous the problem is known as *dimensionality reduction* (*James et al.*, 2013; *Boccignone*, 2017).

Now consider the current data set. We have a collection of eye movement data discretely labeled by the teacher's expertise level or subject matter and the goal is to make inferences about the label using eye movement information. Thus the problem at hand is a binary classification task.

2.4.2 Discriminative vs. Generative Approach

There are two general means of solving the classification problem described in (2.20): discriminative and generative approach. The discriminative approach directly represents the conditional distribution $P(Y \mid X)$ using a parametric model (see (2.21)) of which the parameters θ can be learned from a training set that contains pairings of x_n, y_n (*Bishop and Lasserre*, 2007). The point estimate of θ can be given by maximizing the distribution $P(\theta \mid X, Y)$, then the predictive distribution can be estimated using Eq. (2.23).

$$P(Y \mid X) = \int P(\theta)L(\theta)d\theta \tag{2.21}$$

$$L(\theta) = P(Y \mid X, \theta) = \prod_{n=1}^{N} P(y_n \mid x_n, \theta) \tag{2.22}$$

$$P(\hat{y} \mid \hat{x}, X, Y) \simeq P(\hat{y} \mid \hat{x}, \theta_{MAP}) \tag{2.23}$$

The discriminative approach provides decent predictive performance when the labeled data are abundant, but it may fall short when there are not enough labeled training sets. In this case, an alternative approach that represents both inputs and outputs in parametric model forms may be more useful. This is known as a generative approach since it is capable of generating synthetic data input. The generative approach requires modeling the joint distribution $P(X, Y)$ by modeling $P(X \mid Y)$ and $P(Y)$ first and subsequently calculate $P(Y \mid X)$ using Bayes' Theorem (see Eq. (2.24)) (*Ng and Jordan*, 2002; *Bishop and Lasserre*, 2007).

$$P(Y \mid X) = \frac{P(X \mid Y)P(Y)}{P(X)} \tag{2.24}$$

The discriminative approach includes methods such as logistic regression and Support Vector Machine (SVM), while the generative category contains Naive Bayes, Gaussian Discriminant Analysis (GDA), Hidden Markov Models (HMMs) etc. To put these two approaches in context, consider a simple situation of training a machine to distinguish between cats ($Y = 1$) and dogs ($Y = 0$) based on some physical features such as ear shape, mouth length, etc. Provided with a training set, discriminative algorithms will try to find a decision boundary (a hyperplane) that separates the cats and dogs in terms of the feature space. Then given a new animal, the algorithm will examine on which side of the boundary it falls and make a classification decision accordingly. In the case of the generative approach, the algorithm first builds a model of what cats/dogs look like, and then compares the new animal with each of the models to reach a decision.

Making choices about which approach to adopt requires a series of considerations. First, whether the modeling assumption is satisfied will influence the method choice greatly. For instance, under the circumstance that the input features x are continuous random variables with n dimensions, and output variable y takes values from $0, 1$. *Ng* (2000) has compared two classic methods that model continuous data in two approaches. Logistic regression is a discriminative method that models $P(T \mid X)$ directly as a logistic function of x with parameter(s) θ: $h_\theta(x)$ (2.25). On the other side, Gaussian Discriminant Analysis (GDA) models the data by $P(X, T) = P(T)P(X \mid T)$ in which $P(X \mid T)$ follows a multivariate normal distribution with mean vector $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$ (2.26); the density functions are written as (2.29). Logistic regression and GDA will most often yield different decision boundaries given the same training set, but they are inherently related. *Ng* (2000) showed that this relationship can be shown by rewriting the density function $p(x \mid y = 1)$ as a function of x. The new function can be expressed in the form $p(y = 1 \mid x; \phi; \mu_0, \mu_1, \Sigma) = \frac{1}{1+exp(-\theta^T x)}$, where θ is a function of parameters $\phi, \Sigma, \mu_0, \mu_1$. Compare this function with Eq. (2.25b), *Ng* (2000) pointed out this is the exact form that logistic regression used to model $p(y = 1 \mid x)$, and if $p(x \mid y)$ is a multivariate Gaussian function, then $p(y \mid x)$ necessarily follows a logistic function. But the converse statement is not true, namely, $p(y \mid x)$ being a logistic function does not imply $p(x \mid y)$ follows multivariate Gaussian distribution, thus "this shows that GDA makes a stronger modeling assumption than logistic regression" (*Ng*, 2000; *Ng and Jordan*, 2002). When the input data indeed satisfy Gaussian assumption then GDA is much faster in approaching its asymptotic error and is therefore more efficient and accurate than its discriminative counterparts (*Ng and Jordan*, 2002). On the other hand, logistic regression is more relaxed in terms of modeling assumptions and is robust in most situations.

$$h_\theta(x) = \frac{1}{1 + exp(-\theta^T x)} \tag{2.25a}$$

$$P(y = 1 \mid x; \theta) = h_\theta(x) \tag{2.25b}$$

$$P(y = 0 \mid x; \theta) = 1 - h_\theta(x) \tag{2.25c}$$

$$y \sim Bernoulli(\phi) \tag{2.26}$$

$$x \mid y = 0 \sim \mathcal{N}(\mu_0, \Sigma) \tag{2.27}$$

$$x \mid y = 1 \sim \mathcal{N}(\mu_1, \Sigma) \tag{2.28}$$

$$p(y) = \phi^y (1 - \phi)^{1-y} \tag{2.29}$$

$$p(x \mid y = 0) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp(-\frac{1}{2}(x - \mu_0)^T \Sigma^{-1} (x - \mu_0)) \tag{2.30}$$

$$p(x \mid y = 1) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} exp(-\frac{1}{2}(x - \mu_1)^T \Sigma^{-1} (x - \mu_1)) \tag{2.31}$$

$$|\Sigma|\text{: the determinant of } \Sigma,\ n\text{: number of input features} \tag{2.32}$$

In addition to the consideration of modeling assumptions, the kind of input data also influences method choice. The generative approach typically handles missing data better and is capable of augmenting a small set of labeled data with large quantities of easy-to-acquire unlabeled data. But it is also less direct and slower at making decisions since they are trained to model the joint distribution rather than the direct relationships between class labels and data (*Ulusoy and Bishop*, 2005). In contrast, discriminative approaches are widely used because of their excellent generalization performance when labeled data are plentiful. Another consideration regards the possible incongruence between training and test set. Standard discriminative methods require the training set to contain all possible combinations of feature-label pair, while generative models can handle certain variations (*Bishop and Lasserre*, 2007).

In summary, there is no fixed rule for choosing one approach over another; rather, the decision should be made based on the particular dataset and task. In the next section, I will discuss the process that led to my choice for the current project.

2.4.3 Classification in Eye Movement Research

Both discriminative and generative approaches have been applied in learning from eye movements. In each category, HMM and SVM generally have better performance over other learning algorithms. For example, a growing number of studies have used HMM to model the generative process between certain processing/attentional state and the actual eye movement (*Coutrot et al.*, 2017; *Chuk et al.*, 2017b,a; *Feng*, 2003, 2006; *Simola et al.*, 2008; *Chuk et al.*, 2014), as well as classifying the task a subject is engaged in based on his or her gaze features (*Haji-Abolhassani and Clark*, 2014, 2013; *Kit and Sullivan*, 2016; *Kanan et al.*, 2014; *Borji et al.*, 2015; *Boisvert and Bruce*, 2016). On the other hand, SVM is more widely used to infer a subject's characteristics and identity, such as mental disorders (*Tseng et al.*, 2013; *Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein et al.*, 2011; *Alberdi et al.*, 2016), viewer's gender (*Coutrot et al.*, 2016), age group (*French et al.*, 2017), and level of expertise (*Boccignone et al.*, 2014).

Based on the research question at hand (infer expertise and subject matter), the current paper will adopt SVM for classification purpose.

2.5 Analysis Plan for Mobile Eye Movement Data

2.5.1 Data Transformation

Study 1 and Study 2 will record teacher's eye movements using mobile eye tracking technology. Two types of representations will be used to describe the eye movements: string-based and vector-based representations; each has its unique benefits and limitations.

String-based representation consists of a sequence of characters representing each fixation with a single character. First and foremost, this representation captures the semantic meaning of eye movements, namely what kind of objects were fixated on, thus providing us with highly interpretable data and holding out the possibility of conducting comparisons across classrooms with different layout, student grouping and class content. Second, the string-based representation is also capable of capturing the duration and order information in eye movement sequences. The specific approach of coding fixation duration into string sequences will be discussed in the next section. But at the same time, string-based representation also fails to incorporate important vector properties such as saccade amplitude and direction. Vector representation, on the other hand, preserves the vector properties of scanpath, but is more difficult to interpret. Thus both string and vector representation of eye movement sequences will be used to complement each other.

2.5.1.1 String-based Representation

Semantic Coding Based on our definition of a fixation, a 45 minutes class period contains approximately 4000 fixations. Using the eye tracking video footage, time stamp for each fixation, and room maps with all the students labeled with a numbered ID, trained assistants coded what the teacher was focusing on at the time of fixation. Besides student codes, seven other codes were also used to capture various objects in the classroom (see Table 2.10 for examples). When comparing different classrooms, some labels were combined: different student codes were grouped as *S* and all task-irrelevant objects were combined as *O*. This approach makes the classrooms with different events/students comparable, but unavoidably reduced the information concerning individual students.

Duration Encoding Adapted from the procedure proposed by ScanMatch and SubsMatch 2.0 (*Cristino et al.*, 2010; *Kübler et al.*, 2017), the fixation duration at each fixation location is encoded at a temporal sampling rate of 40 millisecond, thus fixation duration is represented as a repeated sequence of the same label with 40 ms per label (0.75 frame at 30 Hz), for example, a fixation on student for 200 ms can be represented as *SSSSS*. In this way, temporal information can be preserved in a sequence of letters.

Table 2.10: Coding Examples

Semantic Content	Example Label	Explanation
Students	S1∼Snn	Teacher fixating on a particular student
Board	B	For chalkboard/white board used in whole class/large group instruction
Other person	OP	This code is assigned to any student without an ID and any adults (including other teachers, researchers, etc)
Instructional material	IM	Instructional materials (other than the board) that are used as a tool for instruction. This can include any object - and related components - that is intended to help students learn or assist in managing the lesson.
Other objects	OO	Any objects in the classroom that is not used as an instructional material, such as a clock or box of tissue.
Teacher	T	Indicating the teacher is looking at him/herself (for example, look at their hands)
Missing data	M	When the fixation is not captured by eye tracker
Cannot Interpret	X	Unable to judge what the teacher is focusing on

2.5.1.2 Vector-based Representation

The complication of mobile eye movement recording in real life is that the movement is recorded as a point traveling on a two-dimensional (2D) plane, while in actuality the movement unfolds in a three-dimensional (3D) space. The fixations do not fall on a common 2D coordinate system since the teacher's head and body are constantly moving. As discussed in the Methodology chapter, our data bear a limitation of not containing information about head and physical movements, thus the most commonly used eye movement measure—coordinates of fixation locations, is not readily available. In order to adopt the techniques developed under the lab situation in which both subject and display device are static, we need to establish and standardize a common coordinate system first.

Construction of Coordinate System While the exact location of fixation can not be determined uniquely with actual 3D coordinates, there are other plausible ways to represent a scanpath. One way is to reconstruct 3D coordinates from video recordings. Another way is to consider the scanpath not in terms of its position and vector properties in the actual 3D space but rather construct a new space in which the essential properties of the scanpath can be preserved.

The first approach requires algorithms that are designed for automatically extracting a three-dimensional coordinate system directly from the two-dimensional video image. This often requires automatic detection of foreground against background. It is an extremely difficult computational problem that is still at a very primitive stage in related areas such as computer vision and video surveillance. It is beyond this paper's scope to design and implement such algorithms.

The second approach is much more promising given that most fixations in the teaching situation are horizontally distributed because most teaching-related stimuli (e.g., student seats, black boards, teaching equipment, etc.) are positioned so that they are not on top of each other (i.e., we don't have double-decker classrooms). Thus we do not lose important information about identities of fixated objects if we project the three-dimensional space onto a two-dimensional plan viewed from above.

This space-projection idea is inspired by architectural diagrams, especially floor plans, which demonstrate a view from above showing the arrangement of spaces inside the architecture in the same way as a map. Geometrically, the floor plan view is defined as a vertical orthographic projection of an object on to a horizontal plane, with the horizontal plane cutting through the building. In depicting width and length but not height, plans emphasize horizontal arrangements and patterns of function, form, or space (*Ching*, 2015, p. 37). Technically, a floor plan is a horizontal section cut through a building at four feet (one meter and twenty centimeters) above floor level, showing walls, windows and door openings and other features at that level.

Utilizing an actual floor plan based on sketches by observers in the classroom, we constructed a 2D coordinate system with all the objects of interest positioned according to their actual arrangement.

Standardization of the Coordinate System The comparability of different classrooms is the premise of scanpath comparison. Thus before comparing groups of scanpaths, we also need to ensure the data belong to the same coordinate system.

In mobile eye tracking studies, standardization is usually accomplished by pinpointing the location of a relatively static object. For instance, in a speed-stacking task using cups, *Foerster and Schneider* (2013) utilized the position of the bottom cup that was not moved during the task to standardize their coordinate system. Similarly, the anchor could be the location of the shoe during a shoe-tying task, the position of goal location in a navigation task, or the position of a hand during a sandwich-making task.

In the current paper we used the position of the teacher as the origin of the coordinate system. This choice is built on two reasons: teachers were relatively stationary during lecturing; and we are interested in the teacher's perspective. The unit of measure is defined by the closest seating distance between students in a particular classroom, specifically, setting the distance to 1. This arbitrary definition is based on the fact that the seating distance between two students is most often consistent within one classroom. Then the distance between the origin (teacher's position) and other objects can be defined in relation to the shortest distance among students. The distance on the room map was measured using a vector graphic editor (Adobe Illustrator) with a computer-aided design plug-in (CADtools).

Mapping Scheme With the new representation of the classroom, the labeled fixation location can be substituted with (x, y) in the new 2D coordinate system. This results in essentially transforming the eye movement sequences into vectors with dimensions including fixation location (coordinates), saccade direction (vector direction) and amplitude (vector length) etc. Bear in mind that these reconstructed features are not the same as the original oculomotor measures, but since the main goal of this study is not documenting the exact eye movement during teaching (which would not be generalizable to other classrooms), but rather to uncover the variations in teachers' scanpaths as they relate to expertise level and subject area as well as making inferences based on these eye movement sequences, we prefer a transformed sequence of features that retains information about original eye movement patterns.

Yet, there are some complications regarding the mapping of actual eye movements onto points in the newly defined coordinate system that need to be addressed. In the current data set, all the objects relating to instruction are coded with a single label *IM*; similarly, all the objects that are irrelevant are labeled with *OO*. If we are using a single point to represent all the possible fixations, it will greatly decrease the variability of the original scanpath. Thus, in line with the assumption of HMM that the dispersion of fixations follows a two-dimensional Gaussian distribution, we plotted fixations labeled as other objects/person according to a Gaussian distribution (miu=0, $theta$=5) within the range of the whole diagram except for positions occupied by students, board etc. Similarly, fixations labeled as instructional materials were plotted as random points following a uniform distribution within a circle centering around the teacher and with a radius of 1.

Selection of Instructional Units For the sake of comparing different classrooms and teachers, the whole lesson had been segmented into multiple instructional units based on teacher's movement, instructional mode and room arrangement. For the ease of data transformation and group comparison, we used the eye movement when the teacher was relatively stationary and teaching the class as a whole while students remained in their original seats and these instructional units constitutes 53% of the whole class time on average.

2.5.2 Aggregated Similarity Measure

For the vector representation of eye movement, we used MultiMatch (*Dewhurst et al.*, 2012) to get an overall similarity score for scanpath comparison. MultiMatch is a vector-based method that is capable of comparing scanpaths over five dimensions: shape (vector difference), fixation location (position) difference, saccade amplitude (length) difference, saccade direction (angular) difference and fixation duration.

Following a procedure adapted from MultiMatch, the current paper need to first align two scanpaths in order to compare them. The alignment of compared scanpaths is conducted in four steps (also see 2.2.3). First, an eye movement sequence is represented in vector form: start and end point of a vector represents two consecutive fixations, length of the vector denotes the saccade amplitude, and vector direction is essentially saccade direction. The next step is to construct a comparison matrix with elements indicating the pairwise similarity between saccade vectors in two scanpaths. The similarity is represented by vector difference (length of the differential vector), which means each cell in the comparison matrix contains the vector differences between any two saccade vectors, and the vector difference gets smaller when two vectors are similar in magnitude and direction. For the third step, we need to reinterpret this alignment task as a path-finding task. The task of finding a best way to align the two eye movement sequences is essentially to find an optimum path, with a start and end point, through the comparison matrix. Thus the alignment task can be reinterpreted as a path-finding task, which is a classic problem of finding the shortest (minimum cost) path in a graph. Under the framework of graph theory, path-finding is the task of finding the shortest path based on the weights associated with all the edges between two vertices (for definition of terms see Table 2.7). In the case of scanpath comparison, vertices are the elements in a comparison matrix; edges are possible links between comparison matrix elements that are defined by adjacency matrix; and weights are costs associated with each pair-wise comparison (vector difference). In the fourth step, Dijkstra's algorithm (*Dijkstra*, 1959) is used to find the shortest path through the comparison matrix. In a broad term, Dijkstra's algorithm is conducted by repeating two steps: first, visit the unvisited vertex with the smallest known distance from the start vertex; second, for the current vertex, calculate the distance of each unvisited neighbors (vertices that share edges) from the start vertex, if the calculated distance of a vertex is less than the known distance, then update the shortest distance (*Dijkstra*, 1959; *Mehlhorn and Sanders*, 2008). The path found by Dijkstra's algorithm will define how the vectors in two scanpaths can be aligned, namely ensuring that each vector in one scanpath is matched with a vector in the other scanpath (*Jarodzka et al.*, 2010a; *Dewhurst et al.*, 2012).

After alignment, we can then calculate a similarity score for five dimensions: vector shape, saccade amplitude, saccade direction, fixation location and fixation duration

(*Jarodzka et al.*, 2010a; *Dewhurst et al.*, 2012). Each category yields an averaged value over all the aligned pairs of vectors. This value is normalized and then inverted to obtain an interval of 0, 1 with 1 represents perfect match and 0 represents no similarity (see Table 2.8).

2.5.3 Pattern and Classification

2.5.3.1 Point Pattern Analysis

After projecting fixated objects to a set of points on the 2D room map, we not only preserved the vector property of eye movement but also introduced the possibility of conducting point pattern analysis.

In spatial research, point pattern analysis is the evaluation of the pattern, or distribution, of a set of *points* on a surface. Points are defined as the location of an event of interest (*Fotheringham et al.*, 2000). In the current paper, the labeled fixated objects can be considered as points on an isotropic plane with Euclidean distance.

For the purpose of the current paper, I used two point pattern analysis methods to examine the spatial distribution of fixated objects. First, Gaussian kernel estimates of point process intensity were used to evaluate the density of spatial distribution (*Berman and Diggle*, 1989). The bandwidth was determined using Silverman's "rule of thumb" (0.9 times the minimum of the standard deviation and the interquartile range divided by 1.34 times the sample size to the negative one-fifth power) (*Silverman*, 2018, p. 48). This method computes the intensity continuously across the area. I used the kernel smoothed intensity of point pattern to draw a heat map of labeled fixated objects in every classroom.

Another point pattern measure called *standard distance deviation* was used to quantify the extent of spread. Standard distance deviation measures how dispersed a set of points are. It is defined as the standard deviation of the distance of each point from the mean center $(\overline{x}, \overline{y})$ (2.33) (2.34). It is the spatial equivalent of standard deviation and likewise provides a description of the variance of a point set.

$$C = (\overline{x}, \overline{y}) = (\frac{\sum_{n}^{i=1} x_i}{n}, \frac{\sum_{n}^{i=1} y_i}{n}) \tag{2.33}$$

$$d = \sqrt{\frac{\sum_{i=1}^{n}[(x_i - \overline{x})^2 + (y_i - \overline{y})^2]}{n}} \tag{2.34}$$

2.5.3.2 String-based Classification

Finally, for inferring expertise status and subject area from eye movement sequences, I adapted a string-based classification work flow proposed by SubsMatch 2.0 (*Kübler et al.*, 2017). It starts with a string kernel method that segments the eye movement sequence into shorter subsequences and used the frequencies of these subsequences as a feature of similarity. Subsequences are then used to train SVM with a linear kernel.

For the first step, a tool for embedding strings in vector spaces called Sally was used to prepare dataset for classification task (*Rieck et al.*, 2012). Sally implements a generalized *bag-of-words* model in which a text is represented as the set of its words, disregarding the relationship between words but keeping information about their frequency (*Salton et al.*, 1975). In Sally's approach, a string sequence can be characterized by a set of features, such as words or n-grams of bytes, and then each feature is mapped to a high-dimensional vector space whose dimensions are associated with the frequencies of the string features. This association is created using a hash function, where the hash value of each feature defines its dimension. Sally then normalizes the sparse vector that stores the feature frequencies and outputs it in a specified format for further use. Given the characteristics of the current string sequence—50 strings with size of 500–2000 and alphabet size of 5 (S, I, T, O, B)—the length of n-gram is selected at $n = 20$ to account for the repeated pattern in the string.

The second step concerns a binary classification. I used LIBLINEAR, a tool for solving large-scale regularized linear classification (*Fan et al.*, 2008). The L2-regularized L2-loss support vector classification solver for the dual problem was used, and the cost parameter was set to default (1). I have adopted a 10-fold cross-validation approach for the train and test procedure. That is, the training set was equally divided into 10 subsets. At each step one subset is tested using the classifier trained on the remaining 9 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. The cross-validation procedure has been shown to prevent the over fitting problem (*Hsu et al.*, 2003).

CHAPTER III

Study 1: Teacher Expertise and Eye Movement

3.1 Expertise and Eye Movements

Experts are people who show consistently superior performance on representative tasks specific for each domain (*Ericsson and Lehmann*, 1996). What makes them stand out in terms of performance has always been a central research question in learning science. Eye tracking research is one important pathway to understanding the cognitive process of experts.

Interest in how expertise guides eye movement dates back to one of the very first systematic eye movement studies: *Buswell* (1935) compared art students with average viewers when looking at unfamiliar paintings and reported that the experts were more likely to focus on areas that were not centers of interest for other viewers, see also (*Noton and Stark*, 1971b; *Nodine et al.*, 1993; *Zangemeister et al.*, 1995; *Vogt and Magnussen*, 2007; *Humphrey and Underwood*, 2009).

Since then, special gaze behaviors have been found to reflect expert and novice differences: experts notice more than novices do, they hone in on what matters most, and they do so remarkably fast. For example, proficient radiologists can detect an abnormality and recognize it as cancer in a mammogram in under a second (*Kundel et al.*, 2007); and chess masters can report positions of checking pieces while barely moving their eyes (*Reingold et al.*, 2001).

Evidence across a variety of knowledge domains has revealed that experts' and novices' gaze behavior are distinctively different. For example, during the observation of a tumor removal, expert neurosurgeons were found to exhibit larger saccade amplitude and less repetitive fixations after the initial exploration stage (*Eivazi et al.*, 2012; *Kübler et al.*, 2015). Other domains that demonstrated the influence of expertise on eye movement include aviation (*Kasarskis et al.*, 2001; *Schriver et al.*, 2008; *Kennedy et al.*, 2010), sports (*Mann et al.*, 2007; *Memmert et al.*, 2009; *Savelsbergh et al.*, 2002), chess (*Chase and Simon*, 1973; *Reingold et al.*, 2001), examining medical visualizations (*Law et al.*, 2004; *Nodine and Kundel*, 1987; *Pietrzyk et al.*, 2014; *Kübler et al.*, 2015), driving (*Crundall et al.*, 1999; *Crundall and Underwood*, 1998) and reading music (*Kinsler and Carpenter*, 1995; *Waters et al.*, 1997).

A metanalysis by *Gegenfurtner et al.* (2011) that surveyed findings from 296 studies in different expertise domains found that when compared with non-experts (novices and intermediates), experts have shorter fixation durations, more task-relevant fixations as well as fewer task-irrelevant fixations, larger saccade amplitudes and use less time to first fixate relevant information. This effect is moderated by characteristics of visualization (dynamic/static, natural/artificial, with/without text annotation etc.), task properties and expertise domain (*Gegenfurtner et al.*, 2011).

Three theories have been applied to explain the characteristics of expert eye movements. First, the theory of *long-term working memory* proposed that experts store retrieval cues that connect to long-term memory in their working memory, allowing them to retrieve past cognitive processing results in a direct and rapid fashion (*Ericsson and Kintsch*, 1995). This theory proclaims that in order to account for expert performance in domain-specific skilled activities, the limited-capacity assumption of working memory needs to be extended to include working memory based on storage in long-term memory (*Miller*, 1956; *Cowan*, 2010). Given the premise that eye movements reflect cognitive processes underlying task performance and experts acquire shortcuts to retrieve information from long-term memory faster than non-experts, then we can attribute experts' shorter fixation durations to their rapid information retrieval capability (*Gegenfurtner et al.*, 2011).

The second theory to explain experts' eye movement is the *information-reduction* hypothesis (*Haider and Frensch*, 1996, 1999). This theory holds that the speed and quality of expert performance is due to the learned ability to distinguish between relevant and redundant information in order to reduce the amount of task information that needs cognitive processing. If we assume experts are more selective in their use of task information, we can then expect to see experts exhibit shorter fixation duration times, more fixations assigned to task-relevant objects and fewer fixations on irrelevant ones.

Finally, the *perceptual encoding* theory attributes experts' fast and accurate task performance to their encoding of more holistic chunks rather than individual features (*Chase and Simon*, 1973; *Reingold et al.*, 2001). This theory notes that experts' encodings are specific to the task, such that when examining structured, but not random, chess configurations that match their former encodings, experts would make better use of parafoveal processing to extract information from larger and more distant areas, resulting in a greater visual span (*Charness et al.*, 2001). The larger perceptual coverage can explain why experts have fewer fixations and show saccade jumps that span greater areas.

In other words, our working model holds that experts can retrieve information from long-term memory faster, allocate attention more selectively and encode features in larger chunks, all reflected in their eye movement patterns. Compared to steering aircraft, swinging bats, playing chess or performing surgeries, I would argue teaching is no less cognitively challenging. Maneuvering around teaching technologies, guiding students' attention, monitoring their interest and evaluating their response (or lack thereof) and, of course, delivering an effective and interesting lesson all at the same time is indeed a formidable challenge.

Novice teachers who have little experience navigating these elements often fail to notice significant classroom events or miss important "teachable moments" (*König et al.*, 2014; *Seidel and Sturmer*, 2014; *Stockero et al.*, 2017), which might be explained by a phenomenon termed *cognitive tunneling* (*Dirkin*, 1983). Cognitive tunneling describes how the attentional field is narrowed as one engages in an overwhelmingly complex task. A teacher who focuses on only a few students, or a pilot who attends to only a subset of instruments while ignoring or discounting other information would be examples of cognitive tunneling (*Jarmasz et al.*, 2005; *Thomas and Wickens*, 2001). To combat this attention failure, it's essential to understand the movement of the expert's eye during teaching, and the first step will be examining the differences between expert and novice teachers' gaze behavior.

Despite all the revelations about how different expert and novice looking looks, the relationship between expertise and eye movements during teaching remains a relatively unexplored territory (*Stürmer et al.*, 2017; *Cortina et al.*, 2015). The number of studies dedicated to studying teacher's eye movement, especially in natural teaching situations, is quite small. Existing evidence that points in a useful direction includes *Van den Bogert et al.* (2014)'s study that investigated teacher's perception of a video-taped classroom scene. They found that experienced teachers not only process information faster but have more evenly distributed attention across the classroom, leading to better classroom monitoring. Similarly, *Stürmer et al.* (2017) used a mobile eye tracker to record preservice teachers' eye movements while teaching in standardized instructional situations and found a skewed distribution of attention. These studies contribute to framing the potential differences between expert and novice teachers' gaze in the current study, which extends past studies with larger sample size, a naturalistic setting and direct comparison between expert and novice teachers teaching the same students.

The Current Study With the current study, I want to identify the differences between expert and novice teachers' gaze, and also examine the possibility of inferring the expert or novice status of a teacher from the objects they focus on. The current study used records of teachers' eye movement in real classrooms while teachers taught actual lessons, then explored expert and novice teachers' gaze behavior by utilizing statistical methods that match the dynamic property of eye movement.

Since theories of expertise have been generalized to a variety of tasks and domains, it's plausible that expert teachers' eye movement would present similar patterns. Namely, expert teachers might have shorter fixation durations and more fixations, are better at dividing attention among different tasks, and this may be reflected in distinctive scanpath patterns that can be used to infer expertise level.

3.2 Method

3.2.1 Sample

Teachers who participated in the current study were part of the field training element in a teacher preparation program. The existing structure of supervising vs. student teacher naturally constitutes expert-novice pairs. Thus half of the participants were experienced teachers and half of them were their mentees in their final semesters of supervised teaching as part of their certification program. The expert-novice pair therefore taught similar lessons to the identical group of students in the same physical setting.

The current study includes 50 classroom teachers (25 expert-novice pairs) from 25 schools. The schools were located in southeast Michigan covering both affluent and economically challenged neighborhoods. Classes varied in grade level (K–12) and school subject (Math, Science, Literacy, History and Social Studies). Thirty-three of all participating teachers are female and 17 are male.

3.2.2 Apparatus

The main apparatus we used is a Mobile Eye system provided by ASL (Applied Science Laboratories–www.a-s-l.com). This system consists of two parts: a) a head set with an infrared recording camera that keeps track of the wearer's right pupil as well as a small digital camera that takes in the wearer's visual field, and b) a small digital video recorder worn in a fanny pack around the hips. The ASL Mobile Eye records data at 60Hz by interleaving images taken from two camera. Both image streams are recorded on the same digital videotape medium by alternating frames. Therefore, the actual functional sampling of this eye tracker is 30Hz. Data were recorded in a 640 x 480 pixel sensor with a fixed focus video camera.

Based on piloting, we defined eye movement samples as belonging to a fixation if they were moving less than a defined distance (square root of 14 pixels) in either horizontal or vertical dimensions for a minimum of three samples (90 ms).

3.2.3 Procedure

The recording took place during a regular class period selected so as to minimize interference with the regular teaching plan. The running time of a class was generally around 45 minutes. For each teacher in the expert-novice pair, classroom recordings were made on different days with only one of them teaching. Experimenters and observing teachers were also present during class time. The eye tracker can record up to 75 minutes using a battery pack as teachers move freely about their activities. A 5-point system calibration at 5–10-m distance was performed prior to and following the lesson. Two stationary cameras positioned at different angles were also used to provide additional classroom footage, and a third camera tracked the teacher as she taught.

3.3 Analysis Method

The planned analysis work flow will apply to both mobile eye tracking studies (see 2.5 for details). It starts with transforming raw eye movement events into two representations: vector and string representations. First, unlabeled fixations are coded as semantic-based strings, then the strings are transformed to vectors in a newly defined 2-dimensional representation of the classroom. With the string representation, I will be able to apply the method proposed by SubsMatch 2.0, a string kernel method for classifying eye movements based on the frequencies of subsequences in the string (*Kübler et al.*, 2017). SubsMatch 2.0 is built on two tools: Sally (*Rieck et al.*, 2012) for string kernel construction and LIBLINEAR (*Fan et al.*, 2008) for classification, both can be implemented in Unix environment. And with the vector representation, the vector-based MultiMatch algorithm (implemented in MATLAB) that calculates scanpath distances in various dimensions and point pattern analysis method can be utilized to help us understand the overall (dis)similarity between groups (*Dewhurst et al.*, 2012).

3.4 Results

Comparison of Vector Properties The transformed vector representation of raw scan path was used as the input for calculating five dimensions of eye movement properties: vector shape, fixation position, saccade amplitude, saccade direction and fixation duration. By using the MultiMatch algorithm for scan path comparison, each pair of expert-novice teachers received one difference score for each of these five dimensions.

Since every classroom had its own arrangement, student group and subject matter, the difference score between different expert-novice pairs are not comparable, but within-pair comparison is valid since the expert and novice teacher were teaching

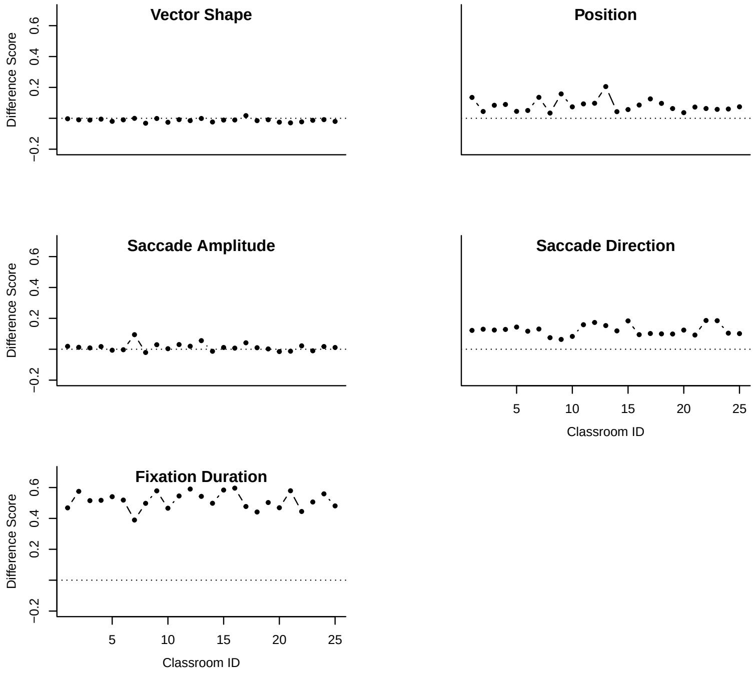

Figure 3.1: Scanpath Comparison Difference Score Compared to Random Baseline. Each classroom ID represents a pair of expert–novice teacher. Dotted line denotes random baseline.

similar content to the same group of students. To evaluate the discrepancies between each expert-novice pair's eye movement properties, expert–novice's scanpath comparison scores was contrasted with expert–random scanpath comparison scores (see Fig. 3.1). The random baseline was computed by sampling with replacement from the original unique coordinates of each expert teacher. In Fig. 3.1, the difference score of twenty-five expert-novice pairs was compared against a random baseline $y = 0$. The differences within a particular expert-novice pair are higher than chance if they are above the random baseline. From Fig. 3.1 we can see that the overall shape of the transformed scanpath is essentially tied with the baseline, saccade amplitude is slightly above baseline for most classrooms, while the other three properties are all above baseline for every expert-novice pair, signaling that the differences between expert and novice teachers' scanpath shapes are not distinguishable from each other, but fixation location, saccade direction and especially fixation duration has a dissimilarity that is above the chance level. Next, I'll explore this dissimilarity in depth.

Distribution of Fixations Firstly, I want to check the distribution of fixated objects within a classroom, and whether experienced and novice teacher demonstrate differences in the density of their visual interests.

As discussed in the analysis method section, the fixated objects can be seen as a set of *points* on a surface that is represented by the room map. Based on this idea, a standard distance deviation can be calculated to represent the degree of spread of experts'/novices' fixations.

Table 3.1 lists the standard distance deviations of all teachers. A numbered ID is assigned to each expert-novice pair, for example, *E01N01* is a pair of expert-novice teachers, with *E01* denoting the expert teacher and *N01* as novice teacher in this pair. Table. 3.1 shows that expert teachers have larger standard distance deviations in 20 out of 25 expert-novice pairs, indicating more variation in the objects they fixated on, more dispersion in visual focus. See Fig. 3.2 for a more visual representation of this trend.

In Fig. 3.2, expert and novice teachers' fixations distributed differently in the same classroom. The fist and third column illustrates the density of expert teacher's fixations, while the second and fourth column presents the fixation locations of the corresponding novice teacher. Brighter color in a particular area indicates denser distribution, and thus showing more visual attention has been directed to this area. We can see that expert teachers' fixations generally cover wider range of the classroom, indicating a more dispersed, more even distribution of visual attention on different objects in the classroom. On the other hand, novice teachers teaching the same classroom seem to have a more skewed distribution, with denser fixations around themselves and on smaller set of objects.

Event-based Comparison As reviewed in first part of this study, past research results have shown experts exhibit different patterns in fixation duration and saccade

Table 3.1: Standard Distance Deviation by Expert–Novice Pair

	Expert	Novice
E01N01	3.841	3.687
E02N02	3.687	1.853
E03N03	1.853	2.740
E04N04	2.740	3.081
E05N05	3.081	2.648
E06N06	2.648	2.536
E07N07	2.536	2.480
E08N08	2.480	2.614
E09N09	2.614	2.402
E10N10	3.494	2.402
E11N11	3.589	3.494
E12N12	3.982	3.589
E13N13	3.982	3.025
E14N14	3.025	1.296
E15N15	1.365	1.296
E16N16	2.437	1.365
E17N17	2.861	2.437
E18N18	2.861	4.416
E19N19	4.579	4.416
E20N20	4.579	2.932
E21N21	3.124	2.932
E22N22	3.124	2.361
E23N23	4.379	2.361
E24N24	4.379	5.756
E25N25	5.756	5.377
Avg.	3.320	2.940

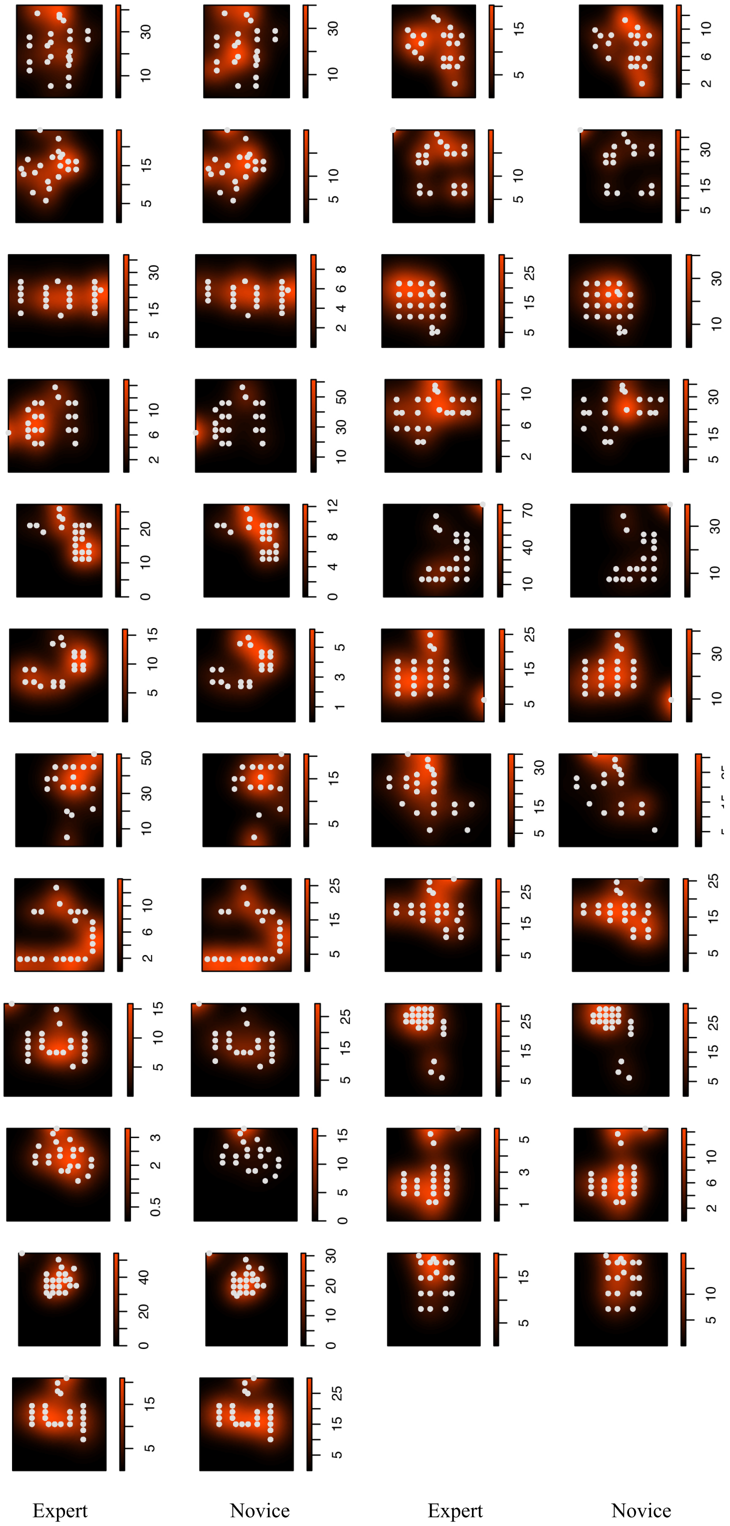

Figure 3.2: Fixated Objects Superimposed on Gaussian Density Plot. Bandwidth selected based on Silverman's rule of thumb. White dots represent the location of fixations.

amplitude. Also, the comparison between expert-novice difference score and random baseline has shown above-chance differences in terms of fixation duration, saccade amplitude and saccade direction within each pair of teachers. Table 3.2 summarizes the value of these three dimensions for all teachers. A two-sample Kolmogorov-Smirnov test was performed within each pair of teachers; the pairs of values that are significantly different at $p < .01$ level have been marked by ** and $p < .05$ marked *. Reading Table 3.2, we can see the general trend is that expert teachers have shorter fixation duration and larger saccade amplitude. But the difference between saccade

Table 3.2: Eye Movement Measures

	Fixation Duration		Saccade Amplitude		Saccade Direction	
	Expert	Novice	Expert	Novice	Expert	Novice
E01N01	0.677	0.562	4.962**	3.983**	0.131	0.125
E02N02	0.804**	1.863**	2.093**	2.736**	0.327	0.326
E03N03	0.843	0.705	3.479	3.716	-0.227	-0.318
E04N04	0.910**	1.053**	3.589**	3.530**	0.211**	-0.111**
E05N05	0.713**	1.773**	3.146**	2.722**	-0.207**	0.045**
E06N06	1.007**	0.929**	3.588	3.479	0.143	0.059
E07N07	0.422**	0.557**	5.167	4.880	0.440	0.398
E08N08	0.769**	0.608**	2.368	1.966	0.493	0.509
E09N09	0.763**	0.859**	3.712**	4.019**	0.163	0.296
E10N10	0.733	0.777	5.449**	4.578**	0.129**	0.048**
E11N11	1.091	1.338	3.423**	2.887**	0.049	0.188
E12N12	0.623**	1.333**	4.353**	4.067**	0.149	0.154
E13N13	0.830**	1.549**	4.726	4.746	0.095	0.257
E14N14	1.285**	1.016**	2.962	2.801	0.027	0.105
E15N15	0.940*	1.289*	2.883	2.976	0.105	0.025
E16N16	1.467	1.691	2.840*	3.351*	0.143	0.138
E17N17	0.992**	0.787**	5.017	5.320	0.336	0.344
E18N18	0.908**	0.570**	4.015	3.752	0.039**	0.374**
E19N19	1.416*	0.900*	2.825	3.242	0.372	0.399
E20N20	0.807*	0.859*	3.513**	2.742**	0.253	0.407
E21N21	1.370	0.692	2.895	2.478	0.128	0.212
E22N22	1.063	0.768	4.290	3.868	0.110	0.126
E23N23	0.779	0.653	2.227**	2.682**	-0.129	-0.082
E24N24	1.783	1.150	2.847**	3.353**	0.217	0.132
E25N25	0.745**	0.627**	3.811**	3.532**	0.173	0.206
Avg.	0.950	0.996	3.607	3.496	0.147	0.174

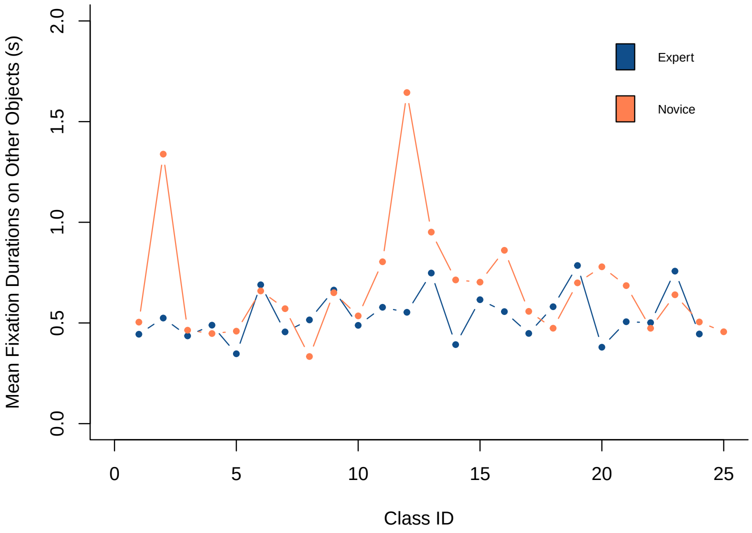

Figure 3.3: Experts' and Novices' Mean Fixation Durations on Task-irrelevant Objects. Each point represents one teacher, class ID is the specific expert-novice pair.

direction is hard to interpret, especially across classrooms. Since the differences between the overall shape of expert and novice's scanpath is not far from chance level, an aggregated measure of direction seems less relevant to the current study. Also the trend with fixation duration is not conclusive since only 12 out of 25 pairs of teacher present this difference, but we need to note that this value counts the overall fixation duration, not the task-relevant duration.

If we break down the fixation duration by task-relevant (students, teacher, board and instructional material) and task-irrelevant (other person or objects) objects, we will have a better insight about the differences between expert and novices' fixation duration (see Figure 3.3 and Figure 3.4). As shown in Figure 3.4, expert teachers have shorter fixation duration time than novice teachers on students, board, instructional material and other objects. By comparing the distribution of two groups of fixation durations, this difference is significant for task-relevant objects ($D = 0.018, p < .001$) and task-irrelevant objects ($D = 0.074, p < .001$).

Similarly, the number of fixations on each category of events present differences between expert and novice teachers' attention distribution. In Figure 3.5 we can see

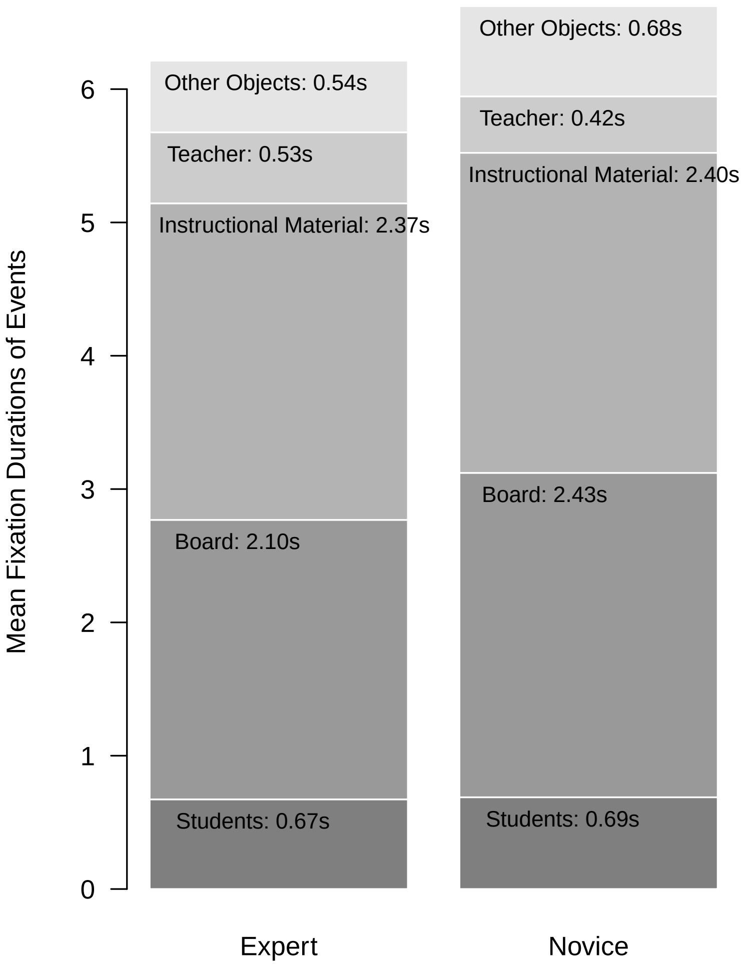

Figure 3.4: Experts' and Novices' Mean Fixation Durations on Different Objects.

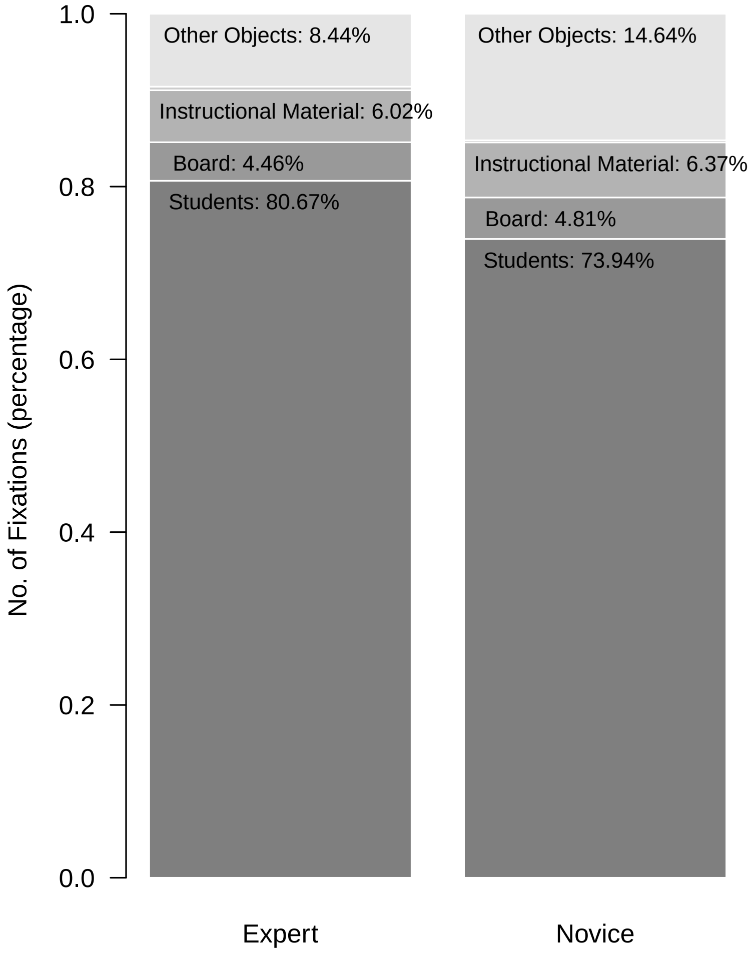

Figure 3.5: Expert and Novice's Number of Fixations on Different Objects. Presented as percentage of the overall number of fixations.

that although both groups would look at students most often, expert teachers have a higher percentage of fixations directed to students, not irrelevant objects, compared to novices.

Classification Finally, the question of whether we can infer teachers' expertise level based purely on their fixated objects is addressed by using a Support Vector Machine (SVM) with a linear kernel. A two-class SVM was trained using a 10-fold cross-validation approach. With a feature length of 20 we reached a classification accuracy of 74%.

This shows with the string representation of scanpath, the model can correctly predict the expertise level of a teacher 74% of the time. The result exceeds the current chance level of 50%. A binomial test showed that this accuracy rate is significantly above chance level 50% at $p < .05$ level.

Breaking down the overall classification accuracy by actual vs. predicted expertise status yields the confusion matrix in Table 3.3. The diagonal of confusion matrix shows the percentage of correct classification (expert as expert, novice as novice), and the off-diagonal represents misclassification rate. The confusion matrix reveals that the model is more likely to correctly label novices compared to expert teachers.

Table 3.3: Confusion Matrix

$n = 50$	Predicted Expert	Predicted Novice
Actual Expert	64%	36%
Actual Novice	16%	84%

After training the model, features with corresponding weights can also be extracted as an indicator of the discriminative power of certain subsequences. Most of these features are difficult to interpret, since the string transformation had long strings of repeated event labels to represent fixation duration. But some of the features did replicate the pattern we found in vector representations, such as the two features show in Listing A.4. Feature 1 has been found to occur in expert teachers' scanpath and feature 2 more often belongs novice teacher. It seems that expert teachers would quickly switch focus between teaching material and students, while novice teachers were likely to fixate on one thing for longer time.

```
Feature 1 : SIIIIISSSSSSSSSSBBBS

Feature 2 : IIIIIIIIIIIIIIIIIISS
```

Listing III.1: Feature Examples

3.5 Discussion

The current study has uncovered differences between expert and novice teachers' eye movement in a real teaching situation. The results showed a similar pattern revealed by past studies on expertise and gaze behavior in other domains.

First, expert teachers exhibit more task-relevant fixations with shorter durations, while novice teachers have more task-irrelevant fixations and longer fixation durations in comparison. The shorter durations can be explained by long-term memory theory, showing that expert teachers might also have convenient retrieval cues connecting working memory and long-term memory, similar with experts from other knowledge domains. And they may have learned to effectively distinguish task-relevant and task-irrelevant events and objects, thus reducing the amount of task information that needs cognitive processing, contributing to a tendency to fixate more on areas related to the current teaching task and avoiding redundant distractions.

Second, the distance between expert teachers' consecutive fixations has larger range, indicating the possible existence of task-specific, holistic encodings that make parafoveal processing more effective.

Finally, both the distribution of fixations and the important classification features demonstrated that experts can attend to a broader array of information, making identifying what's important in classroom situations easier and faster. The wider distribution of fixations may indicate a selective attention allocation.

This pattern of expert viewing can explain why expert teachers often have better *situation awareness* (*Endsley*, 1995; *Endsley and Garland*, 2000). Situation awareness describes the ability to know what is going on around you during a task. It has been defined as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (*Endsley*, 1988, p. 97). By distributing visual attention more widely and by changing focus swiftly, expert teachers may archive better perception of meaningful elements in the classroom, understand the message these elements convey and also be more capable of planning corresponding actions.

The differences between the misclassification rate of expert and novice teachers presented an interesting phenomenon: expert teachers may switch between different gaze pattern, while novice s only know how to look like a novice–focusing on a limited set of objects for long periods of time. This tendency may cause cognitive tunneling that prevents novice teachers from noticing important teaching events in the classroom.

CHAPTER IV

Study 2: Subject Matter and Eye Movement

4.1 Task and Eye Movement

Advocates of knowledge-driven, top-down processes of eye movement generation often start their arguments with *Yarbus* (1967b)'s celebrated study. *Yarbus* (1967b) recorded the drastic discrepancies in people's eye movement pattern when observing the same painting but with different task instructions, such as "estimate the material circumstances of the family in the picture" and "memorize the objects in the picture". Viewers' eyes were directed to the parts of the picture that were informative for the task at hand. For instance, in the age estimation task, viewers were more likely to focus on character's faces, but they tended to fixate on the inanimate objects in the room when the instruction changed to wealth estimation.

Many recent studies have confirmed this result with similar static picture viewing s and more advanced methodology (*Tatler et al.*, 2010; *Ballard and Hayhoe*, 2009; *Kübler et al.*, 2017). For example, *Boisvert and Bruce* (2016) have shown that based on the spatial density of raw fixation positions one can infer the specific task type including object search, saliency viewing and free viewing task.

Evidence from gaze behavior when performing natural tasks also supports the proposition that different eye trajectories emerge as the task requirements vary (*Land et al.*, 1999; *Hayhoe et al.*, 2003; *Land and McLeod*, 2000; *Ballard and Hayhoe*, 2009).

For example, using a sandwich-making task, *Hayhoe et al.* (2003) found that the eye movement pattern during the peanut butter spreading step and jelly spreading step were significantly different. When placing peanut butter on the bread, subjects make anticipatory fixations on the part of the bread where the tip of the knife is going to begin spreading, based on their knowledge that peanut butter often sticks to the knife. In contrast, jelly is less sticky and easier to spread and thus is guided to the bread with a smooth pursuit eye movement. Similarly, *Rothkopf et al.* (2007) demonstrated how fixation location can be driven by task requirements when using an immersive virtual reality environment. Subjects carried out tasks requiring them to either "approach and pick up" or "avoid" certain objects while navigating along a walkway in the virtual environment. They showed considerably different patterns of looking during these two tasks. Subjects' fixations would center on the object when they were instructed to approach the target, while when the task requirement changed to avoiding, their fixations hugged the edge of the object. The visual features of the object remained constant, but as its associated uses change, the fixation distribution also changes. These results provide strong evidence for refuting the notion that eye movements are purely guided by low-level feature conspicuity in the scene. Instead, cognitive control of eye movements is more prevalent in goal-oriented natural tasks.

The common consensus of the top-down approach of eye movement guidance is that human gaze is highly affected by behavioral relevance and learning (*Tatler et al.*, 2011). In Study 1, I examined the relationship between teaching expertise and gaze behavior in the classroom. As a natural extension, I am also interested in how the teaching task can alter teachers' gaze behavior.

Teaching involves a series of smaller tasks, some common across subject matter, but teaching different subjects can be regarded as very different tasks. Field interviews suggest that teachers consider teaching mathematics and literacy to require very different approaches (*Leshem and Markovits*, 2013). Many American teachers hold the view that mathematics is a static body of knowledge, involving a set of axioms and procedures that lead to one correct solution (*Thompson*, 1992; *Nisbet and Warren*, 2000). For example, *Zakaria and Musiran* (2010) investigated beliefs about mathematics among 100 teacher trainees. Most of these preservice teachers believe that mathematics is a formal way of representing reality and that teachers should ensure students acquire a collection of skills and algorithms.

In comparison, teachers often believe that the creativity in language literacy is expressed by the variety of interpretations and the inclusion of students' personal experience. Another interview with elementary school teachers reflected this difference in teacher's belief about math and language literacy: "Mathematics to me is the language of all languages. It is the language of reality. However, while English is a language with an element of emotion—a means by which reality is reflected by words of sentiment—mathematics describes reality in an objective way." (*Leshem and Markovits*, 2013). A teacher also gave the example that "a rose is a rose is a rose" is open to various interpretations, but in mathematics there is only truth: "3 is 3 and 3+4 will always be 7".

The discrepancies in teacher's beliefs about math and literacy very likely lead to adoptions of different teaching strategies and class practices. And these in turn pose different task requirements even when teaching the exact same group of students. Whether teacher's eye movements reflect this variation is the main concern of the current study.

4.2 The Current Study

Teaching a particular subject matter has not yet been considered as a dynamic natural task that poses special task requirements on eye movements. Math and literacy are two subject areas that share great significance in student's development but have often been approached in different ways. If we regard these subjects as two tasks with different properties and requirements, we may expect to see discrepancies reflected in teachers' eye movements. When teacher, students, and classroom environment are all held constant, variations that persist in teachers' eye movements can be interpreted as the influence of task property and demands. This will help us capture real-life teaching practice in a new light and also increase our understanding of top-down process driven gaze behavior.

4.3 Method

4.3.1 Sample

The teachers recruited in this study were from the same area as in Study 1. This study included ten teachers teaching two different subjects, Literacy and Math, with ten classrooms. Students were in grade K–5 and the class size is 15 students on average. Among 10 teachers only one of them is male.

4.3.2 Apparatus

The ASL Mobile Eye system was used for recording teacher's eye movement during teaching. Three external cameras positioned at different angles in the corner of the classroom were also used, one of which followed the teacher as she moved.

The frame rate of the current eye tracking system is 30Hz and the fixation is defined as eye movements that were moving less than a square root of 14 pixels in either horizontal or vertical dimensions for a minimum of three samples (90 ms).

4.3.3 Procedure

Similar to Study 1, the recording took place during a regular class period; the length of a class was about 45 minutes. Each teacher taught both Literacy and Math to the same group of students in the same classroom, but on different days.

4.3.4 Analysis Method

Following the same procedure proposed in Study 1, the raw eye movement sequences were first transformed to string representation denoting the semantic meaning of each fixation. The string coding was carried out by experienced research assistants using frame-by-frame approach. String codes were then mapped on to the room map of the classroom and produced a new vector representation of eye movements with two-dimensional coordinates. With vector representation, teachers' scanpaths were compared using the MultiMatch (*Dewhurst et al.*, 2012) algorithm for alignment and comparison. String representation was recoded with duration information and then used for linear-kernel classification. For details of implementation see 2.5.

4.4 Results

Comparison of Vector Properties The current study includes the same elementary school teachers teaching different subjects (literacy and mathematics) in the same classroom to the same students. For convenience I will will use "classroom" to refer to the unique combination of teacher and subject area, despite the fact the class took part in the same physical environment with unchanging students. For example, teacher *T01* was teaching one literacy class *T01LT* and one math class *T01MA*; *T01LT*/*T01MA* are two different "classrooms".

The scanpaths of the same teacher teaching different subjects were compared along five dimensions: vector shape, fixation position, saccade amplitude, saccade direction and fixation duration. The differences in scanpath of each pair of classrooms were represented as a single (dis)similarity score for each dimension.

To demonstrate the extent of discrepancies between two scanpaths, the (dis)similarity score was compared against a random baseline which is also a score calculated by comparing one scanpath with a randomly generated scanpath. The random scanpath was

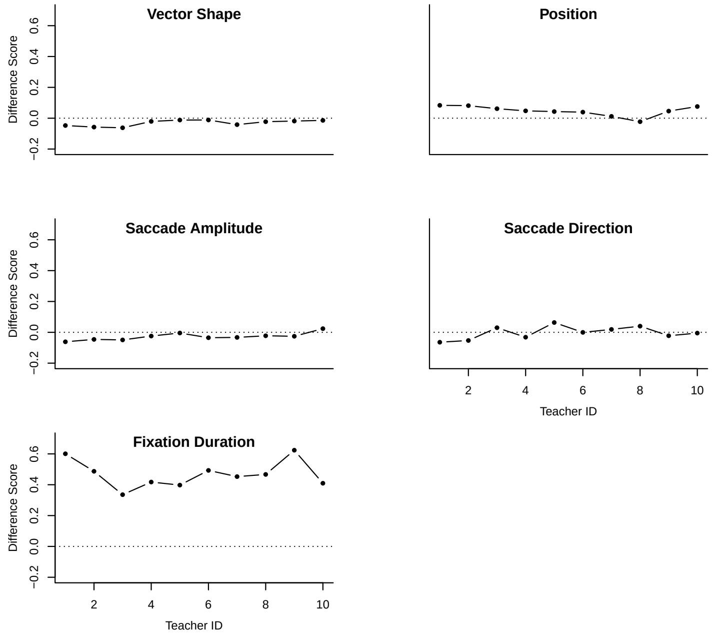

Figure 4.1: Scanpath Comparison Difference Score Compared to Random Baseline. Each teacher ID represents a pair of literacy-math classrooms. Dotted line denotes random baseline.

sampled from all the possible locations and durations of the compared classroom. Fig. 4.1 demonstrates that compared to a random baseline, the distance between each pair of literacy-math classrooms along five dimensions shows different pattern. The overall shape of two scanpaths among all ten teachers was consistently high in similarity; the same pattern also exists in saccade amplitude. These two measures indicate that consistent patterns persist in individual teachers' eye movements across different classrooms.

At the same time, the differences in fixation locations were slightly above chance for almost all teachers. This could imply variation in what objects the teacher visually attended to when teaching different subjects. Similar with what Study 1 found, among five dimensions, fixation duration was the most sensitive measure as to the change of situation: all the teachers exhibit large differences in duration time between the two subjects. Saccade direction, on the other hand, didn't present a consistent pattern.

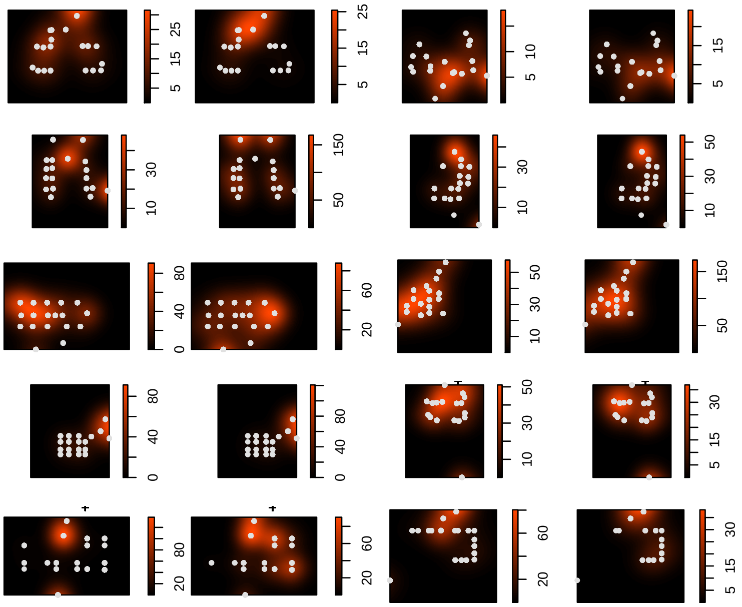

Figure 4.2: Fixated Objects Superimposed on Gaussian Density Plot. White dots represent the location of fixations.

Distribution of Fixations Next, the intensity of fixations on different parts of the classroom was mapped using a Gaussian density plot. In Fig. 4.2, the first and third columns presents the distribution of fixation locations when the teacher was teaching Literacy, and the second and fourth columns are the same teacher teaching Math. As Fig. 4.2 shows, the pattern of fixation distributions were rather consistent for most teachers. Given that the dissimilarity score on fixation locations between math and literacy class was only slightly above chance, this again shows idiosyncratic features of teachers' scanpaths across subject areas when all the fixated objects were taken into account. The similarity across subject matter may also be attributed to the consistent spatial arrangement of the classroom. But since the density plot can not represent the pattern of looking regarding spatially approximate objects, the visual attention directed at teaching-related objects that were clustered around the teacher would not be represented in a distinguishable way in this density plot. For a closer look at the gaze behavior we need to also break down the eye movement measures by object category.

Event-based Comparison As indicated in 4.1, only fixation location and fixation duration presented above-chance variations between math and literacy classrooms. To examine this difference, Table 4.1 summarized the average fixation durations in every classroom with regard to five different categories of objects: students, board, instructional material, teacher and other objects. Tested with a Kolmogorov-Smirnov test, the literacy-math pairs that were significantly different in duration time (seconds) were marked with double asterisk at $p < .01$ level. As shown in the table, there were apparent trends in the length of duration time when students, board and instructional material were considered. The mean fixation durations on students were longer when teaching literacy as compared to math. Teachers also fixated on instructional material longer in literacy classes; these material often included picture books, sample homework and word cards, etc.. On the other hand, fixations were longer toward the board in math classes, possibly because the teacher was demonstrating calculation on the board. There are no consistent pattern as to fixations on teacher themselves and other objects. It is reasonable that duration time on task-irrelevant objects was consistent across math and literacy classes since teachers were teaching in the same environment and were rather familiar with the particular classroom. Number of fixations also a demonstrated a similar pattern (see Table 4.2). Among ten teachers, almost all of them had more fixations on students and instructional material, and less fixations on the board in literacy classroom, while when teaching math they had more fixations on board but less fixations on students and instructional material.

Table 4.1: Fixation Duration by Events

	Students		Board		Instructional Material		Teacher		Other Objects	
	Literacy	Math	Literacy	Math	Literacy	Math	Literacy	Math	Literacy	Math
T01	0.361**	0.215**	0.215**	0.368**	0.327**	0.218**	0.350	0.217	0.304**	0.166**
T02	0.326*	0.287*	0.288	0.421	0.320	0.379	0.310	0.386	0.327	0.309
T03	0.284**	0.142**	0.249	0.221	0.220**	0.153**	0.168	0.221	0.184	0.273
T04	0.248*	0.196*	0.244	0.285	0.158**	0.268**	0.161	0.276	0.208	0.195
T05	0.221	0.222	0.280	0.312	0.254	0.233	0.256	0.247	0.173	0.209
T06	0.304**	0.259**	0.293	0.310	0.251	0.156	0.334	0.282	0.229	0.246
T07	0.240	0.219	0.351	0.314	0.249	0.237	0.300	0.306	0.208	0.270
T08	0.299	0.271	0.236	0.274	0.310	0.287	0.522**	0.279**	0.263	0.277
T09	0.478	0.473	0.312	0.367	0.523**	0.302**	0.299	0.304	0.422**	0.253**
T10	0.228**	0.145**	0.192**	0.289**	0.249**	0.151**	0.252	0.316	0.273	0.179
Avg.	0.310	0.296	0.282	0.313	0.277	0.258	0.270	0.284	0.233	0.245

Table 4.2: Number of Fixations (Percentage of Total)

	Students		Board		Instructional Material	
	Literacy	Math	Literacy	Math	Literacy	Math
T01	80.82%	62.50%	5.28%	9.89%	24.71%	3.96%
T02	74.20%	76.01%	3.69%	1.89%	8.07%	9.20%
T03	76.01%	64.54%	2.92%	29.75%	16.58%	4.39%
T04	93.39%	87.21%	0.06%	3.41%	4.37%	1.34%
T05	92.35%	92.15%	3.86%	4.57%	2.75%	0.52%
T06	89.28%	85.09%	0.84%	9.02%	6.34%	2.01%
T07	85.82%	66.19%	6.05%	14.00%	1.19%	1.67%
T08	72.83%	69.23%	2.63%	15.66%	18.26%	8.49%
T09	79.19%	43.38%	0.07%	2.04%	26.99%	15.05%
T10	69.10%	68.53%	2.34%	29.14%	22.36%	0.66%
Avg.	81.30%	71.48%	2.77%	11.94%	13.16%	4.73%

This trend could have associations with teacher's different beliefs and practice about math and literacy.

Classification In *Kübler et al.* (2017)'s implementation of SubsMatch 2.0, they have reached accuracy of 53% and 59% (25% chance level) when classifying free-viewing and the age-estimation task presented in *Yarbus* (1967b)'s original experiment. They also found that long viewing times influence classification the most. Using a similar string kernel method (same as the one used in Study 1), the L2-regularized L2-loss support vector classification solver for the dual problem was used, the cost parameter was set to the default $c = 1$. The string representations were mapped onto a high-dimension vector space using Sally (*Rieck et al.*, 2012) and then LIBLINEAR (*Fan et al.*, 2008) was used for training and testing the model.

A two-class SVM was trained using a 10-fold cross-validation approach. With a feature length of 20 we reached an average classification accuracy of 75%. Binomial test showed that this accuracy rate is significantly above chance level 50% at $p < .05$ level.

The confusion matrix 4.3 demonstrated that scanpaths of the literacy classroom can be correctly classified 7 out of 10 times, and the prediction for math classroom

Table 4.3: Confusion Matrix

$n = 20$	Predicted Literacy	Predicted Math
Actual Literacy	70%	30%
Actual Math	20%	80%

is even more accurate: 8 out of 10 times. The clear discrepancies in both fixation durations and number of fixations contribute to the discriminability power of this model.

4.5 Discussion

The current study discovered two seemingly contradictory aspects of the eye movement pattern generated by the same teacher teaching different subject matters. One side is that scanpath is a idiosyncratic process; the spatial distribution of fixations of a teacher teaching literacy and math are quite similar despite the different content, format and teaching approach. This result supports the notion that visual exploration is highly idiosyncratic, demonstrated by past findings of the significant similarity of same person's scanpath upon multiple viewings of the same picture (*Noton and Stark*, 1971a; *Andrews and Coppola*, 1999; *Privitera*, 2006; *Rayner et al.*, 2007; *Coutrot et al.*, 2017). This idiosyncratic nature is also demonstrated by comparing the average fixation duration distance in the current study ($M = 0.481$) and the result in Study 1 ($M = 0.402$). Essentially showing that the similarity in fixation duration is higher when the same teacher teaches different contents, as compared to when different teachers teach same content. This replicates past research evidence in the static picture viewing task that duration similarity is higher when comparing the scanpaths of the same person looking at different pictures than when comparing different people on the same picture (*Dewhurst et al.*, 2012). The idiosyncratic nature of scanpaths has been explained by the existence of an internal representation during the initial viewing and the finding that this representation can later serves to guide attention (*Parkhurst and Niebur*, 2003).

The other side of the story presented a considerable difference in how teachers distribute fixations when teaching different subject matters. The current study found that teachers direct their gaze more often at students and instructional material in the literacy class; these fixations were also longer compared to math classes. Checking back with the original footage, we can see that in literacy classes, teachers were telling stories, appraising students' writing and spelling, actively guiding students to generate coherent speech and constantly encouraging students' opinions. In math classrooms, on the other hand, the same teachers were more often using the white board to demonstrate math calculation and concepts; although they were also repeatedly checking student's understanding and connecting the concept with everyday life applications, they were more likely to concentrate on conveying the correct idea as compared to seeking for student's interpretations. This tendency showed up in teachers' longer and more frequent fixations on the board, and fewer fixations on students or instructional material in the math class.

The preferences in fixating on either board or instructional material shows that fixation allocation is often driven by the information-gathering requirements of a particular task, highlighting the role of eye movement as a mechanism of gathering necessary information to accomplish tasks at hand.

The discrepancies in eye movement pattern by subject matter may reflect teachers' different beliefs and practices when approaching these two subject areas. Many teachers believe teaching literacy should be more communicative and conversational, while teaching math is more self-generative in a way that requires long paragraph of explanation.

Regarding teachers' beliefs about mathematics, *Ernest* (1989) has identified three conceptualizations about math: "a dynamic problem-driven view where mathematics is a continually expanding field of human creation; a static, unified body of knowledge where mathematics is viewed as interconnecting structures bound together by logic and meaning; and a bag of tools where mathematics is made up of an accumulation of facts, rules and skills" (also see *Perry et al.*, 1999; *Kuhs and Ball*, 1986).

In relation to these beliefs, two possible teaching approaches may be employed: transmission and constructivist approach. The transmission approach describes the importance of teacher transmitting facts and rules to the students who are expected to memorize and reproduce it. Teachers who take a constructivist approach may consider themselves as facilitators of learning that helps students to construct their own knowledge framework (*Boaler*, 1993; *Nisbet and Warren*, 2000). The results in the current study may reveal that many teachers were acting consistent with the transmission approach for teaching mathematics.

The finding of discrepancies across subject matter does not contradict the first discovery about the idiosyncratic nature of the same individual's scanpath. Teacher's position, and the location of board and instructional material is clustered within a close range compared to the whole classroom space, so the variations in these fixated objects can not be captured by spatial distribution. But this variance can be demonstrated by breaking down the eye movement measures by semantic meaning (what objects are teachers looking at) and temporal attributes (how long do teachers look at the object).

Overall, this study demonstrated both the similarities and differences in the teaching situation where the same teacher approaches two different subject areas.

CHAPTER V

Study 3: Showing the Expert's Eye: Eye Movement Modeling Examples

5.1 Looking from the Expert's Perspective

Study 1 addressed the research question: how does the expert teacher's gaze differ from that of the novice? With an understanding that experts' eye movements can be distinguished from those of novices, a natural follow-up question is whether we can make use of expert's looking to promote learning.

Presenting beginners with videos of an experienced model performing some task is a widely adopted instructional practice in many areas (*Renkl*, 2014; *Van Gog and Rummel*, 2010). As a perfect exemplar of the desired task performance, this kind of instructional video often includes an expert demonstrating the task while verbally explaining each step (*van Gog and Rummel*, 2010). There is growing interest in these *video modeling examples* in both formal and informal educational settings, particularly in online educational settings where their flexibility, accessibility, convenience and individualized learning may have advantages over traditional face-to-face instructions (*Hackbarth*, 1996; *Massy and Zemsky*, 1995; *Fiorella et al.*, 2017). For instance, students may refer to online courses on how to solve derivatives or watch a video for learning how to play the ukulele.

To increase the effectiveness of such video modeling examples, it may be beneficial to consider not only demonstrating the expert's external modeling behavior but also their inner processes as reflected in eye movements.

In a comprehensive review of eye tracking research on visual expertise in chess and medicine, *Reingold et al.* (2001) has concluded that the expert's way of encoding domain-specific patterns can be reflected in their eye movements. And experts' eye movements often reflect their tacit knowledge about a particular task that even they themselves are unaware of, or unable to verbalize.

van Gog et al. (2009) pioneered the concept of *eye movement modeling examples* (EMME), which includes not only the recording of a model's demonstration but also the overlay of the model's eye movement as he performs the task in order to guide learner's attention in information-dense, dynamic and complex situations. *Litchfield et al.* (2010) found that by using eye movements as cues for guiding the viewer's attention, the performance of pulmonary node detection in X-rays improved. *Jarodzka et al.* (2013) further showed that the inclusion of eye movement in video modeling examples can facilitate learning. Using a task of classifying the swimming modes of reef fish, a task common in the domain of marine zoology, *Jarodzka et al.* (2013) showed that EMME improved students' performance through refining their visual search strategies and interpretation of relevant information. They argued that the benefit of EMME stems from the externalization of the otherwise inaccessible attention allocation process that may "synchronize" students' attention with the model. This benefit is particularly relevant when the modeled task requires highly strategized visual attention allocation in dynamic and realistic situations, such as reading a dashboard when flying a plane, interpreting medical visualizations during surgery, or reading music score (*Jarodzka et al.*, 2009, 2010b, 2017).

Besides the tacit knowledge and attention guidance possibly embedded in expert gaze, eye movement video may also be beneficial because it is a unique way of showing the expert's point of view. As *Fiorella et al.* (2017) correctly points out, most of the research on video modeling examples concern the characteristics of the model, including gender, age and expertise level (*Hoogerheide et al.*, 2016), but much less about an important design feature of instructional video, namely, the perspective of the video. Showing the eye movement of the model also provides the viewer with the egocentric perspective of this expert model, since seeing another's eye movements is inherently a first-person perspective experience.

Perspective, also called *centricity*, *frame of reference*, or *point of view*, is used to refer to the extent to which a human observer's viewpoint is removed from the nominal viewpoint with respect to the agent in the media (*McCormick et al.*, 1998; *Salzman et al.*, 1999). Defined by the perceived view point, perspective is often specified as either first-person perspective (1PP or egocentric perspective) or third-person perspective (3PP or exocentric perspective). 1PP is an immersed perspective, presenting the observer the viewpoint of being inside the video. This view is thus also referred to as an "inside-out" view. In contrast, 3PP is an "outside-in" or bird's-eye perspective that gives the observer the vantage point of overlooking all or a large portion of the environment in the video. Within this view, viewers are able to see movements of the agent, as if they are looking at the agent from above, or from the side, or from some other external viewpoints (*Milgram and Colquhoun Jr*, 1999).

First-person perspective contributes greatly to the sense of immersion. Immersion is the subjective impression that one is participating in a comprehensive, realistic experience (*Sadowski and Stanney*, 2002; *Lessiter et al.*, 2001). Immersion in a digital context involves the willing suspension of disbelief, to voluntarily believe the virtual world to be true and real. The design of immersive experiences that induce this suspension of disbelief usually draws on sensory, actional, and symbolic factors (*Dede*, 2009).

Salzman et al. (1999) found that the exocentric and the egocentric perspectives may have different strengths for learning. They claimed that a major advantage of egocentric perspectives is that they enable the viewer's actional immersion and motivation through embodied, concrete learning, whereas exocentric perspectives foster more abstract, symbolic insights gained from distancing oneself from the context. In a similar light, studies have shown that video perspective affects instruction in which objects are being manipulated (*Castro-Alonso et al.*, 2015; *Carbonneau and Marley*, 2015; *Fiorella et al.*, 2017). For example, using a circuit assembly task, *Fiorella et al.* (2017) showed that students watching first-person perspective modeling videos perform significantly better compared to the control group that watched the third-person perspective version.

In sum, eye movement modeling examples may provide tacit expert knowledge that is otherwise unavailable, visual cues for attention guidance, a feeling of immersiveness and an embodied learning experience. Based on these properties, the current study would expect to see a performance improvement in learner's speed and accuracy of completing a perceptual task after watching an eye movement modeling example video.

5.2 The Current Study

This study concerns whether showing an expert's visual perspective in a video modeling example can improve learners' performance in a spatial puzzle task.

In addition to the presence or absence of eye movement information, another relevant variable often seen in first-perspective modeling videos is the inclusion of the model's hand. The effect of showing/not showing the model's hand in a motor task has been shown in past studies using object manipulative tasks, including LEGO constructions (*Castro-Alonso et al.*, 2015) and knot tying (*Marcus et al.*, 2013).

The influence of showing the model's hand is associated with the concept of *perihand space*, which describes the space near our and other people's hands. The hand is the theater of most of our interactions with the external world, reflecting not only spatial perception and action but also their integration. Many active learning activities happen in this space: reading, writing and manipulating objects, to name a few.

Substantial neurological and behavioral evidence exists in favor of the existence of a special hand-centered representation of space, i.e., perihand space (*Brozzoli et al.*, 2014). The neurophysiological findings define a set of at least four distinctive areas that relate to perihand space perception: premotor area 6, parietal areas 7b and ventral section of the intraparietal sulcus, and the putamen (*Hyvärinen and Poranen*, 1974; *Avillac et al.*, 2005; *Duhamel et al.*, 1998). This means the human brain takes the hand as the reference point for coding the space "immediately outward the hand and follows it, staying anchored to this reference when the hand changes location" (*Brozzoli et al.*, 2014, p. 125).

Following this line of reasoning, showing the hand of the model might elicit a stronger sense of immersiveness, providing a more embodied learning experience. Thus showing a model's hand or not is introduced as another factor that might influence the efficacy of the current modeling video.

In the present study, I examined the effect of perspective information embedded in an example modeling video which might promote problem solving performance in a perceptual task. The instructional video varied along two dimensions: perihand space information and eye movement information. These two aspects of video will be compared in terms of their influence on learning outcomes.

5.3 Method

5.3.1 Participants

Two hundred college students were recruited from an undergraduate subject pool. Participants were roughly gender balanced and had no past experience with the perceptual task used.

5.3.2 Study Design

The participants were randomly divided into 4 groups of equal size. The example modeling video varied on two between-subject factors: the presence of eye movement information (with/without eye movement information), and the presence of hand information (with/without model's hand).

The learning performance was evaluated using number of errors and completion time before and after watching the modeling video. The time factor serves as a within-subject factor. Participants also took a short survey about the effectiveness of the modeling video after completing the task.

5.3.3 Material

The current study used a perceptual task that includes direct manipulation of objects. Perceptual tasks rely heavily on visual search strategies and interpretations of perceptual information, an example could be reading X-ray scans (*Chi*, 2006; *Jarodzka et al.*, 2013). The task used in this study was a maze game displayed on a mobile device. This task was chosen based on several considerations: a) it requires specific order of inspection, namely a visual scan strategy is necessary for fast and accurate completion; b) it involves manually controlling an avatar walking through a maze; c) the difficulty level can be easily controlled.

The objective of this maze task is to find a path out of the maze without doubling back, while avoiding certain obstacles at the same time. A necessary strategy for this game is to narrow the set of possible paths by identifying the exit of the maze. Because only one exit exists in the maze, the anticipated path will be incorrect if it results in a new endpoint (no exit). This strategy was reflected in the eye movement information in the form of prolonged fixation on corners of the maze, and moving the eye backwards from the exit to the start point.

The model's eye movement during the task was collected using the Tobii model T60 eye tracker. For the hand factor, the modeling videos that present hand information were shot using the SMI Eye tracking glasses mobile eye tracker without actually recording eye movement, essentially using it as a head-mounted camera. The videos without hand information were simply the screen recordings of the mobile device.

5.3.4 Measurement

Participants' task performance was evaluated by their completion time in seconds and number of errors. An error was defined as a backward motion along the path, so each time the participants change their routes and went backward would be counted as one error.

A short survey including four evaluative questions was also used. Participants were instructed to rate the following questions V.1 from 1 (Completely Disagree) to 5 (Completely Agree) based on their experience.

1. How helpful was the instructional video in helping you to solve the maze?
2. I felt that I was seeing things from the model's point of view when I was watching the video.
3. I could tell what the model's strategy for solving the maze was when I was watching the video.
4. How engaging did you find the video to be?

Listing V.1: Evaluative Questions

5.3.5 Procedure

After a short introduction about the study and the task, participants were asked to solve the easiest level as a practice to familiarize themselves with the task. Then they would solve two levels independently as a pretest. After that, participants watched one of the four instructional videos that ran about 2 minutes during which they were instructed to put themselves in the mental shoes of the model and pay attention to the eye movement information/hand information.

After the modeling video, the participants were provided with another two problems that were similar in difficulty with the pretests. Participants' performances were recorded and compared with the pretest. Their subjective feeling of whether the gaze information and hand information helped their learning and whether watching the video made them feel like looking through the model's point of view were also recorded.

5.4 Results

Changes in Completion Time and Number of Errors Ten participants were removed from the current analysis because of incomplete pretest and six more were excluded as they did not finish the post-test. Participants' task performances were rated by two scores: completion time and number of errors. The changes in completion time and number of errors before and after watching the modeling video were calculated by averaging between two pretest and two post-test levels and then deducting the pretest score from the post-test. A negative change score indicates a decrease in completion time or number of errors.

Table 5.1 summarizes the average change score in four conditions. We can see that across all conditions, participants performed better in terms of a shorter completion time and fewer errors after watching the video modeling examples. Also it is clear that the best combination is the no hand-with eye movement condition, in which both change scores are larger than other three video conditions.

Table 5.1: Summary Statistics of Four Conditions

Hand Information	Eye Movement Information	*n*	Changes in Completion Time (sec)		Changes in Number of Errors	
			M	*SD*	*M*	*SD*
Yes	Yes	47	-6.085	13.156	-3.723	1.930
No	Yes	46	-11.413	30.222	-4.435	3.557
Yes	No	50	-6.660	11.849	-3.180	0.962
No	No	53	-6.868	13.776	-3.792	2.222

Table 5.2: Analysis of Variance: Completion Time

	Df	Sum Sq	Mean Sq	*F*	*p*
Time	1	9.100	9.140	2.043	0.155
Hand:Time	1	0.003	0.003	0.022	0.883
Eye:Time	1	0.030	0.030	0.238	0.627
Hand:Eye:Time	1	0.001	0.001	0.010	0.919
Residuals	180	22.507	0.125		

Error term: subject by time

Table 5.3: Analysis of Variance: Number of Errors

	Df	Sum Sq	Mean Sq	*F*	*p*
Time	1	27.200	27.174	7.709	0.006 **
Hand:Time	1	3.100	3.141	0.891	0.346
Eye:Time	1	19.600	19.612	5.564	0.019 *
Hand:Eye:Time	1	0.600	0.611	0.173	0.678
Residuals	180	634.500	3.525		

Error term: subject by time

Performing MANOVA on task performance scores yields different results for hand and eye movement information, and for two performance measures. Table 5.2 and 5.3 shows that the main effect of eye information was only significant when the performance was measured by changes in number of errors, and hand information didn't have a significant effect on both changes in completion time and changes in number of errors. This result indicates that immediately after watching the video modeling examples, learners didn't present sizable speed improvement when solving new mazes, but their rate of error dropped significantly when the video included the model's eye movement. This may suggest that task speed is more resistant to changes and requires deliberate practice to improve, but knowledge about strategy may be grasped quite quickly from a model's demonstration and eye movement and induce a decrease in errors.

Survey Response The short survey provided us with more insights about the effect of eye movement information. Firstly, learners didn't think differently about the general helpfulness of the video modeling example with or without the inclusion of model's eye movement, they consistently agreed the video was helpful. But the information of eye movement did make them felt like "seeing things from the model's point of view" ($t(198) = 2.038$, $p < .05$) and noticed "the model's strategy for solving the maze" ($t(198) = 2.2964$, $p < .05$). And finally, learners consistently rated it as engaging with or without eye movement information.

There were also interesting gender differences in how learners felt about the eye movement information (see Table 5.4). Across all questions and eye movement conditions, female learners rated higher on the experience of watching modeling videos. Compared to their male counterparts, female learners found the video to be more helpful and engaging, as well as easier to extract perspective and strategy information. This result could be reflect a more positive attitude on the part of the female participants, or it might indicate that they were more attuned to the idea of learning from modeling videos.

5.5 Discussion

This study suggested that learners can benefit from observing not only an expert's demonstration but also their eye movements. The guidance provided by the model's

Table 5.4: Mean Survey Response Score

	With Eye Movement			Without Eye Movement		
	Female	Male	Avg.	Female	Male	Avg.
Q_Helpful	3.023*	2.431*	2.727	2.854	2.604	2.729
Q_FPV	2.409	2.232	2.321*	2.220*	1.904*	2.062*
Q_Strategy	2.422	2.224	2.323*	2.118	1.887	2.002*
Q_Engaging	3.222*	2.724*	2.973	2.924	2.921	2.922

eye movements is effective likely because the task requires special strategies for visual inspection that are difficult to convey verbally. The benefits of showing the expert's eye were not only reflected in learner's performance improvement as measured by a decrease of errors, but also their subjective feeling of being able to see from expert's perspective, as well as being able to grasp the essential strategy just from watching expert's eye movement.

Eye movement modeling examples have been shown to be effective in learning to classify dynamic motion pattern (*Jarodzka et al.*, 2013) and text-picture processing strategy (*Mason et al.*, 2015). The result of the current study also indicate that expert's eye movement information is useful when visual inspection strategy is involved.

Compared to eye movement information, the inclusion of expert's hands didn't have a significant effect in changing learner's performance. This might be due to the fact that the information and strategy provided by eye movement and hand movement partially overlapped, causing a redundancy effect. The current perceptual task includes a planning step and an action step. During planning, the expert looked at the maze and explored possible solutions without moving the hands, while in the action step the expert would carry out the solution and move the hand around the maze. At this stage, the eye movement obviously followed that movement, thus the information provided by hand and eye is indistinguishable. This redundancy effect is also suggested by *van Gog et al.* (2009), with a procedural problem-solving task where they found that model's eye movement accompanied with verbal explanation actually had a negative effect on subsequent transfer task performance. They reasoned that verbal explanation or eye movement alone might have been sufficient for guiding learner's attention, presenting both had caused the redundancy effect that hinder learning (*van Gog et al.*, 2009; *van Marlen et al.*, 2016; *Mayer and Johnson*, 2008). Similarly, when hand movement and eye movement are presenting the same guidance, the combination of two may be detrimental, not beneficial for learning (*van Marlen et al.*, 2016; *Mayer and Johnson*, 2008).

This study also revealed differences in how different genders feel about the eye movement modeling examples. Female learners consistently rate the learning experience more positive compared to their male counterparts, and they agreed that expert's point of view and strategy were clear. The results about hand information and gender differences are reminders that there are nuanced considerations about applying eye movement modeling examples for instructional purposes.

Eye movement information opens another way of presenting the expert's perspective, revealing experts' thinking and reasoning processes and increasing learners' embodied learning experience. This study has demonstrated the unique potential of making the expert's eye movements explicit. And with the fast development of visual and interactive technology such as head-mounted virtual reality devices and eye-capture functionality in mobile device, it's expected that eye movement information will play an even more important role in shaping modern instruction and learning.

CHAPTER VI

Discussion

6.1 Main Findings

Learning what we can discover from teachers' eye movements is the primary motivation of the current paper. The current endeavor is initiated with the belief that capturing teachers' eye movements can deepen our understanding about teaching by addressing the following questions: 1) how do teachers distribute attention in the course of instruction; 2) what is the influence of expertise level on teachers' eye movements; 3) are there patterns in how experienced/novice teacher move their eyes; 4) are teacher's eye movement patterns consistent across different subject areas; and 5) how can we apply expert's vision in implementing instruction?

Study 1: Teacher Expertise and Eye Movement Findings from Study 1 reveal that expert and novice teachers' eye movements are considerably different: expert teachers have more task relevant fixations with shorter durations, larger visual span and wider fixation allocations. Expert teachers, just like experts from other areas such as chess and medicine, have distinctive eye movement patterns that may indicate their special domain-specific knowledge organizations. The efficient organization of large amount of information such as *chunking* in chess is one basis for experts' superior performance (*Chase and Simon*, 1973; *Jarodzka et al.*, 2017). And this perceptual encoding has been found in the way the eye moves. *Reingold et al.* (2001) demonstrated that experts do not fix their eyes on one single chess figure, but look rather in between figures (*Charness et al.*, 2001; *Reingold et al.*, 2001; *Reingold, E. M., & Sheridan*, 2011). The result in Study 1 presented a similar pattern, indicating possible chunking in teacher's perceptual processes.

This idea of expert teachers possessing a distinctive structure of knowledge organization is also proposed by researchers in educational science. Expert teacher's content, pedagogical and pedagogical content knowledge can not only be characterized by the sizable amount of knowledge, but especially by having a superior organization of that knowledge (*Livingston and Borko*, 1989; *Krauss et al.*, 2008; *Leinhardt and Greeno*, 1986; *Lachner et al.*, 2016). As teachers gain experience in teaching, they tend to organize their knowledge around encountered cases and experiences, which may result in more elaborated and coherently organized knowledge structures (*Krauss et al.*, 2008; *Putnam*, 1987).

In sum, Study 1 has shown distinctive eye movement patterns of expert teachers in real teaching situations that support the idea that expertise in teaching is associated with a distinctive pattern of looking.

Study 2: Subject Matter and Eye Movement Study 2 examined the relationship between subject matter and teachers' eye movements using the same methodological procedures used in Study 1. Both similarities and discrepancies were found in the same teacher's eye movements when teaching different subjects. The overall scanpath shape and spatial distribution of fixations for a particular teacher were consistent across math and literacy classrooms, but teachers' fixation allocation on certain objects changes according to the subject area, indicating a difference in teachers' belief and practice regarding math and literacy class.

Past studies suggested that mathematics is often regarded as a rigorous subject that requires knowledge transfer to be precise and meticulous. During this process, teachers serve as transmitters of facts and rules, and students are expected to follow certain steps to reach conclusion. On the other hand, literacy is seen as a more inclusive subject in which learning occurs when students are actively involved in constructing meaning for themselves through activities and discussions. In this view, teachers are seen as facilitators of learning rather than transmitters of facts (*Nisbet and Warren*, 2000).

This study found that teachers direct their gaze more at students and instruction material during literacy class while more visual attention is put upon the board when teaching math. This discrepancy may reflect differences in teachers' beliefs and the resulting teaching approach for math and literacy.

Study 3: Showing Expert's Eye: Eye Movement Modeling Examples The application of using expert's eye movement in instruction was explored in Study 3. Using a perceptual task that requires special visual inspection strategy, Study 3 revealed the benefit of showing an expert's eye movements alongside the traditional demonstration in video modeling examples: the number of errors significantly decreased after watching an eye movement modeling example video. Displaying the model's hands also decreased learners' errors by enhancing the embodied learning experience. Learners reported positively about the feeling of immersion and ease of extracting visual inspection strategies. Female learners showed higher ratings on these scales compared to male learners.

Findings from Study 3 favor the application of expert eye movements when teaching perceptual tasks that require specific but implicit visual inspection strategies.

6.2 Considerations and Future Directions

Bottom-up Influences in the Classroom The classroom is a great testbed for understanding visual cognition in the context of complicated natural behavior. The current paper adopts solely the perspective of top-down approach about eye movement generation and guidance; in other words I focused on how personal characteristics (expertise) and task requirements (subject matter) shape teachers' eye movements, but ignored the alternative view of the bottom-up approach: how visually salient objects and events can affect teachers' eye movements. Eye movement is the product of the combination of all the motor and perceptual activities involved during a task. Therefore in principle, all such factors should be taken into account when modeling human eye movements.

Saliency-based approaches are often regarded as unsuitable for understanding eye movements in natural settings (*Renninger et al.*, 2007; *Tatler et al.*, 2011), based on the reasoning that most daily tasks have clear means-end relationship that call upon the viewer's past experience to guide eye movement. But teaching is a different situation in which the means-end relations are usually more abstract and less structured. Visual attention during teaching is motivated by multiple co-occurring processes: biomechanical factors, conspicuities of objects in the environment, task parameters and personal characteristics (*Tatler and Vincent*, 2009b). Thus it might be important to consider the importance of visual conspicuities in the classroom. The saliency of visual features in the classroom may be more relevant for novice teachers as they haven't acquired the right way to navigate in the complex environment with demanding cognitive tasks. Students' sudden movements (standing up, raising a hand, misbehaving) and salient task-irrelevant objects may catch novice's attention more easily and hold their fixations for a longer time.

If we assume experienced teachers are knowledge-driven, then novices may be more stimulus-driven. To test this hypothesis, a future study that models visual conspicuity in a three-dimensional environment is necessary. A possible next step for the current paper might be creating a saliency map of certain classrooms, and if this map can readily predict novice teachers' eye movements but does less well for expert teacher, then we can establish connections between expertise level and bottom-up processes.

Three-dimensional Eye Movement Representation The analysis of eye tracking data in the current paper is based on a transformed 2D graphic representation of the environment. This representation is not an exact replica of the classroom, but rather a simplified representation in proportional to the true arrangement. Although this transformation has enabled us to compare the vector properties of eye movement sequences, it certainly requires improvement as the raw location information were never recorded. Note especially that the transformation was unable to provide information about the vertical dimensions of forms and spaces, while the information in the third dimension is certainly also relevant for our research purposes.

Therefore, the next step will require both a 3D coordinate system and tracking of teachers' physical movement to map fixation locations in the environment accurately.

Another limitation concerns the selection of scenes. In order to use a consistent room map for constructing a new 2D coordinating system, the current paper had separated the whole class into different segments according to teacher's position and class activity. Only the segments when the teacher was relatively stationary and delivering a lecture were used in the analysis, the more dynamic group activities were unavoidably excluded from the current paper. This approach does not indicate the insignificance of group activities in understanding teacher's eye movement, quite the contrary, the interactions between teacher and students and the resulting gaze behavior could be very interesting. Equipped with 3D coordinates and movement tracking data, the eye movement when teacher and students are both moving will be available for analysis and in general do a better job for capturing the dynamics of classroom.

Cultural Variations of Expert Teacher's Eye Movement The classroom is a microcosm of society, and culture shapes teaching practice. *Hofstede* (1986) has examined classroom dynamics in 50 countries based on the 4-D model of cultural differences. This model describes societies along four dimensions: Individualism versus Collectivism, large versus small Power Distance, strong versus weak Uncertainty Avoidance, and Masculinity versus Femininity (*Hofstede*, 1986).

In the first dimension, Individualist cultures assume that people primarily look after their own or their immediate family's interest. Collectivist cultures assume everyone belongs to certain groups, the group affiliation status not only protects the individual's interest but also demands the individual's loyalty. When reflected in teacher-student interaction, individualist classrooms encourage individual students to speak up in larger groups, accepting open confrontation and conflicts. Because a collectivist classroom strives to maintain harmony at all times, individual students will only speak up in class when called upon personally or in small groups.

The second dimension, Power Distance defines the extent to which the less powerful minorities accept inequality in power and consider it as normal. In a society that is large in power distance, teacher-centered education prevails: students expect teachers to initiate communication, lead the path of learning, and never criticize what the teacher says. By comparison, small power distance gives students permission to speak spontaneously in class, contradict teachers, and construct their own personal learning experience.

The third dimension, Uncertainty Avoidance is a cultural characteristic that defines the extent to which people strive to avoid unstructured or unpredictable situations by maintaining strict codes of behavior accompanied by belief in absolute truths. In strong uncertainty avoidance classrooms, students prefer structured learning situations with clear objectives, teachers are expected to know all the answers, and only reward students with correct solutions. In contrast, weak uncertainty avoidance classrooms have a loose structure, teachers are approachable as they use plain language and need not know every answer. Students are praised for unconventional, creative answers.

Finally, Masculinity and Femininity describes the social roles attributed to men and women in each society. "Feminine" classrooms are designed for average students, students are allowed to learning according to their intrinsic interest and their social skills are highly rated. On the other side, "masculine" classrooms use the best students as the norm, reward students' academic performance and students' motivation is more related to practical goals and incentives.

With the stated differences in classroom dynamics in relation to the particular societal values it stem from, it is expected that teacher's gaze may look very different in various cultures. For example, *Yamamoto and Imai-Matsumura* (2013) attempted to replicate results from *Van Gog et al.* (2005) but found that when watching a classroom video, Japanese expert teachers were not more aware of student's misbehaviors than novice teachers. There were also no significant differences in fixation duration or the time to the first fixation on a target student. In another comparison study, *McIntyre et al.* (2017) found that teachers from UK had shown greater efficiency in attentional gaze (gaze used for information-seeking) than their East Asian counterparts, whereas teachers in Hong Kong displayed greater efficiency in communicative gaze (gaze used for information-giving).

Given the fact that most eye tracking and expertise studies were conducted among US or European samples (including the current paper), the issue of generalization has to be addressed with more varied cultural samples in future studies.

6.3 Implications

Implications for Teacher Education. Teaching is not an easy task. *Berliner* (2001) has poignantly compared classrooms with high-pressure "nuclear power plants, medical emergency rooms and air traffic controls" due to the fact that teaching is a process operating at parallel levels that require fast decision-making (*Haider et al.*, 2005), complex maneuvers (*Chassy and Gobet*, 2011) and superior memory capacities (*Saariluoma*, 1991; *McIntyre et al.*, 2017). Expert teachers show various advantages in terms of conveying the subject matter, managing the classroom (*Kunter et al.*, 2013; *Wolff et al.*, 2017), assessing students' performance (*Ruiz-Primo and Furtak*, 2007) and applying core-practice of teaching (*Forzani*, 2014). All these advantages of expertise have to build on how teachers look at the classroom.

Lachner et al. (2016) prosed a model of teachers' cognition in relation to teaching practices and aspects of the situational context in which teacher's *professional vision* is the building block of the rest of processes. Teachers' professional vision is the ability to notice and interpret relevant features of classroom situations (*Goodwin*, 1994; *Van Es and Sherin*, 2002; *Seidel and Sturmer*, 2014). This ability constitutes an important part of teaching expertise.

Novice teachers have been found to struggle with the complexity of the classroom environment and to apply what they have learned to the context of the real teaching situation (*Stokking et al.*, 2003). They are often not able to direct their attention to relevant objects, events and situations that impact student learning (*Star and Strickland*, 2008). Novice teachers don't naturally know what to look at and where to find it in the classroom, they must learn not only the locations at which relevant information can be found, but how to distribute fixations and when to do so. And current teacher training practice does not include preparing novice teachers with such knowledge and skill. Therefore, one of the key aims of teacher training should be fostering the acquisition of professional vision (*Seidel and Sturmer*, 2014). Most past research on teachers' professional vision used qualitative approaches or indirect assessment of teachers' ability to notice and reason about classroom situations depicted on video; thus understanding how expert and novice teachers actually look in real teaching situations has tremendous importance, both in increasing our understanding about instruction and in facilitating teacher training.

With the findings in Study 1, the next question will be whether this kind of expert vision can be trained, and what kinds of practices are likely to promote its development. Here again, research in the acquisition of perceptual skills in other domains provides clues to the kind of viewing experiences that are likely to be effective. In the domain of sports, several studies have looked at the extent to which perceptual training can lead to increases in performance. *Harle and Vickers* (2001) showed that training basketball players in where and how to look prior to the shooting action led to increases in free throw performance. Another example used video-based perceptual training based on model's perspective and found improvements in tennis playing by intermediate level players (*Farrow et al.*, 1998; *Williams et al.*, 2002). Similar results were also found in other athletic domains, including cricket (*Müller et al.*, 2006).

Building on the findings of this paper, it is likely that we can design better teacher education materials that incorporate teacher-perspective videos. For example, this might involve using expert teacher's eye movement modeling videos to teach novices how to notice and manage misbehaviors in the classroom, overcome their limited situation awareness, avoid the so called "cognitive tunneling", and in general fostering the acquisition of professional vision and teaching expertise.

Implications for Modern Instruction The practical significance of these studies lies in the fact that eye tracking technology has became increasingly accessible and economic friendly for the general public. Low-cost, open-source eye trackers aiming at a mass market are starting to appear. The possibility of making traditional instructional videos more appealing and effective by adding another layer of eye movement information is also promising. In an era of fast technology development, understanding both the process and application of learning from eye movements will be even more relevant for educators and learners alike.

APPENDIX A

APPENDIX A

Configurations

A.1 Configurations of Sally

```
# Configuration of input
input = {
    input_format = "lines";
    lines_regex = "^(\\+|-)?[0-9]+";
};

# Configuration of feature extraction
features = {
    # Length of n-grams.
    ngram_len = 20;
    # Granularity of n-grams.
    granularity = "bytes";
    # Delimiters for tokens.
    token_delim = "";
    # Embedding mode for vectors. Supported types "cnt", "bin", "
       tfidf"
    vect_embed = "cnt";
    # Normalization mode for vectors. Supported types "l1", "l2", "
       none".
    vect_norm = "l2";
};


# Configuration of output
output = {
    # Format of output. Supported types: "libsvm", "text", "matlab"
    output_format = "libsvm";
};
```

Listing A.1: Configurtion File

```
# R produced one .txt file for each classroom, concatenate all .txt
   files
awk 'FNR==1{print␣""}1' *.txt > all.txt
# Commandlines to run Sally: use configuration file testbyte.cfg, read
    all.txt and output all.libsvm
sally -c testbyte.cfg all.txt all.libsvm
```

Listing A.2: Terminal Commands

A.2 Configurations of LIBLINEAR

```
train -v 10 -c 1 all.libsvm # 10-fold cross validation with default
   solver, yields overall classrification accuracy
```

Listing A.3: Terminal Commands

```
[test_label, test_inst] = libsvmread(’all.libsvm’)
model = train(test_label, test_inst, ’-c␣1’)
[predict_label, accuracy, dec_values] = predict(test_label, test_inst,
    model)
```

Listing A.4: MATLAB Syntax

BIBLIOGRAPHY

BIBLIOGRAPHY

Ahissar, M., and S. Hochstein (1997), Task difficulty and the specificity of perceptual learning, *Nature*, *387*(6631), 401.

Alberdi, A., A. Aztiria, and A. Basarab (2016), On the early diagnosis of Alzheimer's Disease from multimodal signals: A survey, *Artificial Intelligence in Medicine*, *71*, 1–29, doi:10.1016/j.artmed.2016.06.003.

Andrews, T. J., and D. M. Coppola (1999), Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments, *Vision research*, *39*(17), 2947–2953.

Andrienko, G., N. Andrienko, M. Burch, and D. Weiskopf (2012), Visual analytics methodology for eye movement studies, *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2889–2898, doi:10.1109/TVCG.2012.276.

Avillac, M., S. Deneve, E. Olivier, A. Pouget, and J.-R. Duhamel (2005), Reference frames for representing visual and tactile locations in parietal cortex, *Nature neuroscience*, *8*(7), 941.

Bacon, W. F., and H. E. Egeth (1994), Overriding stimulus-driven attentional capture, *Perception & psychophysics*, *55*(5), 485–496.

Ball, D. L. (2000), Bridging practices: Intertwining content and pedagogy in teaching and learning to teach, *Journal of teacher education*, *51*(3), 241–247.

Ballard, D. H., and M. M. Hayhoe (2009), Modelling the role of task in the control of gaze, *Visual Cognition*, *17*(6-7), 1185–1204, doi:10.1080/13506280902978477.

Ballard, D. H., M. M. Hayhoe, and J. B. Pelz (1995), Memory representations in natural tasks, *Journal of Cognitive Neuroscience*, *7*(1), 66–80.

Barnes, G. R. (2011), Ocular pursuit movements, in *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, chap. 7, pp. 115–132, Oxford University Press.

Baum, L. E., and T. Petrie (1966), Statistical inference for probabilistic functions of finite state markov chains, *The annals of mathematical statistics*, *37*(6), 1554–1563.

Bergen, J. R., and B. Julesz (1983), Parallel versus serial processing in rapid pattern discrimination, *Nature*, *303*(5919), 696–698.

Berliner, D. C. (2001), Learning about and learning from expert teachers, *International journal of educational research*, *35*(5), 463–482.

Berman, M., and P. Diggle (1989), Estimating weighted integrals of the second-order intensity of a spatial point process, *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 81–92.

Betz, T. (2010), Investigating task-dependent top-down effects on overt visual attention, *Journal of Vision*, *10*(3), 1–14, doi:10.1167/10.3.15.

Bishop, C., and J. Lasserre (2007), Generative or discriminative? getting the best of both worlds, *Bayesian Statistics*, *8*, 3–24.

Boaler, J. (1993), The role of contexts in the mathematics classroom: Do they make mathematics more" real"?, *For the learning of mathematics*, *13*(2), 12–17.

Boccignone, G. (2017), Advanced statistical methods for eye movement analysis and modeling: a gentle introduction, *arXiv:1506.07194 [physics, q-bio]*.

Boccignone, G., M. Ferraro, and S. Crespi (2014), Detecting expert's eye using a multiple-kernel Relevance Vector Machine, *Journal of eye . . .*, *7*(2), 1–15, doi:10.16910/jemr.7.2.3.

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

Bondy, J. A., U. S. R. Murty, et al. (1976), *Graph theory with applications*, vol. 290, Citeseer.

Borji, A., and L. Itti (2013), State-of-the-art in visual attention modeling, *IEEE transactions on pattern analysis and machine intelligence*, *35*(1), 185–207.

Borji, A., A. Lennartz, and M. Pomplun (2015), What do eyes reveal about the mind?. Algorithmic inference of search targets from fixations, *Neurocomputing*, *149*(PB), 788–799, doi:10.1016/j.neucom.2014.07.055.

Bouma, H. (1970), Interaction effects in parafoveal letter recognition, *Nature*, *226*(5241), 177–178.

Brozzoli, C., H. H. Ehrsson, and A. Farnè (2014), Multisensory representation of the space near the hand: from perception to action and interindividual interactions, *The Neuroscientist*, *20*(2), 122–135.

Buswell, G. T. (1935), *How people look at pictures*, University of Chicago Press Chicago.

Caldara, R., and S. Miellet (2011), iMap: A novel method for statistical fixation mapping of eye movement data, *Behavior Research Methods*, *43*(3), 864–878, doi:10.3758/s13428-011-0092-x.

Carbonneau, K. J., and S. C. Marley (2015), Instructional guidance and realism of manipulatives influence preschool children's mathematics learning, *The Journal of Experimental Education*, *83*(4), 495–513.

Carpenter, R. H. (1988), *Movements of the Eyes, 2nd Rev*, Pion Limited.

Castro-Alonso, J. C., P. Ayres, and F. Paas (2015), Animations showing lego manipulative tasks: Three potential moderators of effectiveness, *Computers & Education*, *85*, 1–13.

Charness, N., E. M. Reingold, M. Pomplun, and D. M. Stampe (2001), The perceptual aspect of skilled performance in chess: Evidence from eye movements, *Memory & cognition*, *29*(8), 1146–1152.

Chase, W. G., and H. A. Simon (1973), The mind's eye in chess, in *Visual information processing*, pp. 215–281, Elsevier.

Chassy, P., and F. Gobet (2011), Measuring chess experts' single-use sequence knowledge: an archival study of departure from theoreticalopenings, *PLoS One*, *6*(11), e26,692.

Chi, M. (2006), Laboratory methods for assessing experts' and novices' knowledge, *Cambridge Handbook of Expertise and Expert Performance*, pp. 167–184.

Ching, F. D. (2015), *Architectural graphics*, John Wiley & Sons.

Chuk, T., A. B. Chan, and J. H. Hsiao (2014), Understanding eye movements in face recognition using hidden Markov models, *Journal of vision*, *14*(2014), 1–14, doi:10.1167/14.11.8.doi.

Chuk, T., A. B. Chan, and J. H. Hsiao (2017a), Is having similar eye movement patterns during face learning and recognition beneficial for recognition performance? Evidence from hidden Markov modeling, *Vision Research*, *141*, 204–216, doi:10.1016/j.visres.2017.03.010.

Chuk, T., K. Crookes, W. G. Hayward, A. B. Chan, and J. H. Hsiao (2017b), Hidden Markov model analysis reveals the advantage of analytic eye movement patterns in face recognition across cultures, *Cognition*, *169*, 102–117, doi:10.1016/j.cognition.2017.08.003.

Cinlar, E. (2013), *Introduction to stochastic processes*, Courier Corporation.

Coco, M. I. (2009), The statistical challenge of scan-path analysis, *Proceedings - 2009 2nd Conference on Human System Interactions, HSI '09*, pp. 372–375, doi:10.1109/HSI.2009.5091008.

Cortina, K. S., K. F. Miller, R. McKenzie, and A. Epstein (2015), Where Low and High Inference Data Converge: Validation of CLASS Assessment of Mathematics Instruction Using Mobile Eye Tracking with Expert and Novice Teachers, *International Journal of Science and Mathematics Education*, *13*(2), 389–403, doi: 10.1007/s10763-014-9610-5.

Coutrot, A., N. Binetti, C. Harrison, I. Mareschal, and A. Johnston (2016), Face exploration dynamics differentiate men and women, *Journal of Vision*, *16*(14), 16, doi:10.1167/16.14.16.

Coutrot, A., J. H. Hsiao, and A. B. Chan (2017), Scanpath modeling and classification with hidden Markov models, *Behavior Research Methods*, pp. 1–18, doi:10.3758/s13428-017-0876-8.

Cowan, N. (2010), The magical mystery four: How is working memory capacity limited, and why?, *Current directions in psychological science*, *19*(1), 51–57.

Cristino, F., S. Math??t, J. Theeuwes, and I. D. Gilchrist (2010), ScanMatch: A novel method for comparing fixation sequences, *Behavior Research Methods*, *42*(3), 692–700, doi:10.3758/BRM.42.3.692.

Crundall, D., G. Underwood, and P. Chapman (1999), Driving experience and the functional field of view, *Perception*, *28*(9), 1075–1087.

Crundall, D. E., and G. Underwood (1998), Effects of experience and processing demands on visual information acquisition in drivers, *Ergonomics*, *41*(4), 448–458.

DeAngelus, M., and J. B. Pelz (2009), Top-down control of eye movements: Yarbus revisited, *Visual Cognition*, *17*(6-7), 790–811, doi:10.1080/13506280902793843.

Dede, C. (2009), Immersive interfaces for engagement and learning, *science*, *323*(5910), 66–69.

Derrick J. Parkhurst, E. N. (2004), Texture contrast attracts overt visual attention in natural scenes, *European Journal of Neuroscience*, *19*, 783–789, doi:10.1111/j.1460-9568.2003.03183.x.

Dewhurst, R., M. Nyström, H. Jarodzka, T. Foulsham, R. Johansson, and K. Holmqvist (2012), It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach, *Behavior Research Methods*, *44*(4), 1079–1100, doi:10.3758/s13428-012-0212-2.

Dijkstra, E. W. (1959), A note on two problems in connexion with graphs, *Numerische mathematik*, *1*(1), 269–271.

Dirkin, G. (1983), Cognitive tunneling: Use of visual information under stress, *Perceptual and Motor Skills*, *56*(1), 191–198.

Duchowski, A. T. (2007a), *Neuroscience and Psychology*, chap. 17, pp. 207–240, Springer.

Duchowski, A. T. (2007b), *Taxonomy and Models of Eye Movements*, chap. 4, pp. 41–48, Springer.

Duhamel, J.-R., C. L. Colby, and M. E. Goldberg (1998), Ventral intraparietal area of the macaque: congruent visual and somatic response properties, *Journal of neurophysiology*, *79*(1), 126–136.

Eivazi, S., R. Bednarik, M. Tukiainen, M. von und zu Fraunberg, V. Leinonen, and J. E. Jääskeläinen (2012), Gaze behaviour of expert and novice microneurosurgeons differs during observations of tumor removal recordings, in *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 377–380, ACM.

Ellis, S. R., and J. D. Smith (1985), Patterns of statistical dependency in visual scanning, *In Eye Movements and Human Information Processing*, *9*, 221–238.

Endsley, M. R. (1988), Design and evaluation for situation awareness enhancement, in *Proceedings of the Human Factors Society annual meeting*, vol. 32, pp. 97–101, SAGE Publications Sage CA: Los Angeles, CA.

Endsley, M. R. (1995), Toward a theory of situation awareness in dynamic systems, *Human factors*, *37*(1), 32–64.

Endsley, M. R., and D. J. Garland (2000), *Situation awareness analysis and measurement*, CRC Press.

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

Ericsson, K. A., and A. C. Lehmann (1996), Expert and exceptional performance: Evidence of maximal adaptation to task constraints, *Annual review of psychology*, *47*(1), 273–305.

Ernest, P. (1989), The impact of beliefs on the teaching of mathematics, *Mathematics teaching: The state of the art*, *249*, 254.

Fan, R.-E., K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin (2008), Liblinear: A library for large linear classification, *Journal of machine learning research*, *9*(Aug), 1871–1874.

Farrow, D., P. Chivers, C. Hardingham, and S. Sachse (1998), The effect of video-based perceptual training on the tennis return of serve., *International Journal of Sport Psychology*.

Feng, G. (2003), From eye movement to cognition: Toward a general framework of inference comment on Liechty et al., 2003, *Psychometrika*, *68*(4), 551–556, doi: 10.1007/BF02295610.

Feng, G. (2006), Eye movements as time-series random variables: A stochastic model of eye movement control in reading, *Cognitive Systems Research*, *7*(1), 70–95, doi: 10.1016/j.cogsys.2005.07.004.

Findlay, J. M., and I. D. Gilchrist (2008), Active Vision: The Psychology of Looking and Seeing, in *Active Vision: The Psychology of Looking and Seeing*, pp. 1–240, Oxford University Press, doi:10.1093/acprof:oso/9780198524793.001.0001.

Fiorella, L., T. V. Gog, V. Hoogerheide, R. E. Mayer, L. Fiorella, and R. E. Mayer (2017), It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task., *Journal of Educational Psychology*, *109*, 653.

Flexer, A., P. Sykacek, I. Rezek, and G. Dorffner (2000), Using hidden markov models to build an automatic, continuous and probabilistic sleep stager, in *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on*, vol. 3, pp. 627–631, IEEE.

Flexer, A., G. Gruber, and G. Dorffner (2005), A reliable probabilistic sleep stager based on a single eeg signal, *Artificial Intelligence in Medicine*, *33*(3), 199–207.

Foerster, R., and W. Schneider (2013), Functionally sequenced scanpath similarity method (FuncSim): Comparing and evaluating scanpath similarity based on a task's inherent sequence of functional, *Journal of Eye Movement Research*, *6*(5), 1–22, doi:10.16910/jemr.6.5.4.

Forzani, F. M. (2014), Understanding core practices and practice-based teacher education: Learning from the past, *Journal of Teacher Education*, *65*(4), 357–368.

Fotheringham, A. S., C. Brunsdon, and M. Charlton (2000), *Quantitative geography: perspectives on spatial data analysis*, Sage.

French, R. M., Y. Glady, and J. P. Thibaut (2017), An evaluation of scanpath-comparison and machine-learning classification algorithms used to study the dynamics of analogy making, *Behavior Research Methods*, *49*(4), 1291–1302, doi: 10.3758/s13428-016-0788-z.

Friedman, J., T. Hastie, and R. Tibshirani (2001), *The elements of statistical learning*, vol. 1, Springer series in statistics New York.

Frintrop, S., E. Rome, and H. I. Christensen (2010), Computational Visual Attention Systems and their Cognitive Foundations: A Survey, *ACM Journal Name*, *7*(1), 1–39, doi:10.1145/1658349.1658355.

Fuster, J. (2004), Upper processing stages of the perceptionâ action cycle, *Trends in Cognitive Sciences*, *8*(4), 143–145, doi:10.1016/j.tics.2004.02.004.

Gegenfurtner, A., E. Lehtinen, and R. Säljö (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, doi:10.1007/s10648-011-9174-7.

Ghahramani, Z. (2001), An introduction to hidden Markov models and Bayesian networks., *International Journal of Pattern Recognition and Artificial Intelligence*, *15*(01), 9–42, doi:10.1142/S0218001401000836.

Gilchrist, I. (2011), Saccades, in *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, chap. 5, pp. 85–94, Oxford University Press.

Goodwin, C. (1994), Professional vision, *American anthropologist*, *96*(3), 606–633.

Greene, M. R., T. Liu, and J. M. Wolfe (2012), Reconsidering Yarbus: A failure to predict observers' task from eye movement patterns, *Vision Research*, *62*, 1–8, doi:10.1016/j.visres.2012.03.019.

Hackbarth, S. (1996), *The educational technology handbook: a comprehensive guide: process and products for learning*, Educational Technology.

Haider, H., and P. A. Frensch (1996), The role of information reduction in skill acquisition, *Cognitive Psychology*, *30*(3), 304–337, doi:10.1006/cogp.1996.0009.

Haider, H., and P. A. Frensch (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, doi:10.1037/0278-7393.25.1.172.

Haider, H., P. A. Frensch, and D. Joram (2005), Are strategy shifts caused by data-driven processes or by voluntary processes?, *Consciousness and Cognition*, *14*(3), 495–519.

Haji-Abolhassani, A., and J. J. Clark (2013), A computational model for task inference in visual search, *Journal of Vision*, *13*(3), 29–29, doi:10.1167/13.3.29.

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

Harle, S. K., and J. N. Vickers (2001), Training quiet eye improves accuracy in the basketball free throw, *The Sport Psychologist*, *15*(3), 289–305.

Hayhoe, M. M., A. Shrivastava, R. Mruczek, and J. B. Pelz (2003), Visual memory and motor planning in a natural task, *Journal of vision*, *3*(1), 6–6.

Heinke, D., and G. Humphreys (2005), Computational models of visual selective attention: A review, *Connectionist models in cognitive psychology*, *1*(Part 4), 273–312.

Henderson, J. M. (2011), Eye movements and scene perception, in *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, chap. 33, pp. 593–606, Oxford University Press.

Henderson, J. M., and A. Hollingworth (1999), High-level scene perception, *Annual review of psychology*, *50*(1), 243–271.

Hillstrom, A. P., and S. Yantis (1994), Visual motion and attentional capture, *Perception & psychophysics*, *55*(4), 399–411.

Hofstede, G. (1986), Cultural differences in teaching and learning, *International Journal of Intercultural Relations*, *10*(3), 301–320, doi:10.1016/0147-1767(86)90015-5.

Hollingworth, A., G. Schrock, and J. M. Henderson (2001), Change detection in the flicker paradigm: The role of fixation position within the scene, *Memory & Cognition*, *29*(2), 296–304.

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer (2011a), *Eye-tracker Hardware and its Properties*, chap. 2, pp. 21–24, OUP Oxford.

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer (2011b), *Attention Maps-Scientific Tools or Fancy Visualizations?*, chap. 7, pp. 231–252, OUP Oxford.

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer (2011c), *Scanpaths-Theoretical Principles and Practical Application*, chap. 8, pp. 253–285, OUP Oxford.

Hoogerheide, V., M. van Wermeskerken, S. M. Loyens, and T. van Gog (2016), Learning from video modeling examples: Content kept equal, adults are more effective models than peers, *Learning and Instruction*, *44*, 22–30.

Hsu, C.-W., C.-C. Chang, C.-J. Lin, et al. (2003), A practical guide to support vector classification.

Humphrey, K., and G. Underwood (2009), Domain knowledge moderates the influence of visual saliency in scene recognition, *British Journal of Psychology*, *100*(2), 377–398.

Hyvärinen, J., and A. Poranen (1974), Function of the parietal associative area 7 as revealed from cellular discharges in alert monkeys, *Brain*, *97*(4), 673–692.

Itti, L. (2005), Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes, *Visual Cognition*, *12*(6), 1093–1123, doi:10.1080/13506280444000661.

Itti, L., and C. Koch (2000), A saliency-based search mechanism for overt and covert shifts of visual attention, *Vision research*, *40*(10), 1489–1506.

Itti, L., and C. Koch (2001), Computational modelling of visual attention., *Nature reviews. Neuroscience*, *2*(3), 194–203, doi:10.1038/35058500.

Itti, L., C. Koch, and E. Niebur (1998), A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013), *An introduction to statistical learning*, vol. 112, Springer.

Jarmasz, J., C. M. Herdman, and K. R. Johannsdottir (2005), Object-based attention and cognitive tunneling, *Journal of Experimental Psychology: Applied*, *11*(1), 3–12, doi:10.1037/1076-898X.11.1.3.

Jarodzka, H., K. Scheiter, P. Gerjets, T. van Gog, and M. Dorr (2009), How to Convey Perceptual Skills by Displaying Experts' Gaze Data, *Cogsci*, pp. 2920–2925, doi:10.1016/j.learninstruc.2009.02.019.Lindsey.

Jarodzka, H., K. Holmqvist, and M. Nyström (2010a), A vector-based, multidimensional scanpath similarity measure, *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications - ETRA '10*, *1*(212), 211, doi:10.1145/1743666.1743718.

Jarodzka, H., K. Scheiter, P. Gerjets, and T. van Gog (2010b), In the eyes of the beholder: How experts and novices interpret dynamic stimuli, *Learning and Instruction*, *20*(2), 146–154, doi:10.1016/j.learninstruc.2009.02.019.

Jarodzka, H., T. Van Gog, M. Dorr, K. Scheiter, and P. Gerjets (2013), Learning to see: Guiding students' attention via a Model's eye movements fosters learning, *Learning and Instruction*, *25*, 62–70, doi:10.1016/j.learninstruc.2012.11.004.

Jarodzka, H., K. Holmqvist, and H. Gruber (2017), Eye tracking in educational science : Theoretical frameworks and research agendas, *Journal of Eye Movement Research*, *10*(1), 1–18, doi:10.16910/jemr.10.1.3.

Jovancevic, J., B. Sullivan, and M. Hayhoe (2006), Control of attention and gaze in complex environments, *Journal of Vision*, *6*(12), 9, doi:10.1167/6.12.9.

Julesz, B. (1984), A brief outline of the texton theory of human vision, *Trends in Neurosciences*, *7*(2), 41–45.

Kanan, C., N. a. Ray, D. N. F. Bseiso, J. H. Hsiao, and G. W. Cottrell (2014), Predicting an observer's task using multi-fixation pattern analysis, *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14*, pp. 287–290, doi:10.1145/2578153.2578208.

Kasarskis, P., J. Stehwien, J. Hickox, A. Aretz, and C. Wickens (2001), Comparison of expert and novice scan behaviors during vfr flight, in *Proceedings of the 11th International Symposium on Aviation Psychology*, vol. 6, Citeseer.

Keane, T. P., N. D. Cahill, and J. B. Pelz (2014), Eye-movement sequence statistics and hypothesis-testing with classical recurrence analysis, in *Proceedings of the Symposium on Eye Tracking Research and Applications - ETRA '14*, January, pp. 143–150, doi:10.1145/2578153.2578174.

Kennedy, Q., J. L. Taylor, G. Reade, and J. A. Yesavage (2010), Age and expertise effects in aviation decision making and flight control in a flight simulator, *Aviation, space, and environmental medicine*, *81*(5), 489–497.

Kinsler, V., and R. Carpenter (1995), Saccadic eye movements while reading music, *Vision Research*, *35*(10), 1447–1458.

Kit, D., and B. Sullivan (2016), Classifying mobile eye tracking data with hidden Markov models, *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct - MobileHCI '16*, pp. 1037–1040, doi:10.1145/2957265.2965014.

Koch, C., and S. Ullman (1985), Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry, *Human neurobiology*, *4*(4), 219–227, doi:10.1007/978-94-009-3833-5_5.

Koch, C., and S. Ullman (1987), Shifts in selective visual attention: towards the underlying neural circuitry, in *Matters of intelligence*, chap. 4, pp. 115–141, Springer.

König, J., S. Blömeke, P. Klein, U. Suhl, A. Busse, and G. Kaiser (2014), Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach, *Teaching and Teacher Education*, *38*, 76–88, doi:10.1016/j.tate.2013.11.004.

Kowler, E. (2011), Eye movements: The past 25years, *Vision Research*, *51*(13), 1457–1483, doi:10.1016/j.visres.2010.12.014.

Krauss, S., M. Brunner, M. Kunter, J. Baumert, W. Blum, M. Neubrand, and A. Jordan (2008), Pedagogical content knowledge and content knowledge of secondary mathematics teachers., *Journal of Educational Psychology*, *100*(3), 716.

Kübler, T. C., S. Eivazi, and E. Kasneci (2015), Automated visual scanpath analysis reveals the expertise level of micro-neurosurgeons, *MICCAI'15 Workshop on Interventional Microscopy*, pp. 1–8.

Kübler, T. C., C. Rothe, U. Schiefer, W. Rosenstiel, and E. Kasneci (2017), SubsMatch 2.0: Scanpath comparison and classification based on subsequence frequencies, *Behavior Research Methods*, *49*(3), 1048–1064, doi:10.3758/s13428-016-0765-6.

Kuhs, T. M., and D. L. Ball (1986), Approaches to teaching mathematics: Mapping the domains of knowledge, skills, and dispositions, *East Lansing: Michigan State University, Center on Teacher Education*.

Kundel, H. L., C. F. Nodine, E. F. Conant, and S. P. Weinstein (2007), Holistic component of image perception in mammogram interpretation: gaze-tracking study, *Radiology*, *242*(2), 396–402.

Kunter, M., U. Klusmann, J. Baumert, D. Richter, T. Voss, and A. Hachfeld (2013), Professional competence of teachers: effects on instructional quality and student development., *Journal of Educational Psychology*, *105*(3), 805.

Lachner, A., H. Jarodzka, and M. Nückles (2016), What makes an expert teacher? Investigating teachers' professional vision and discourse abilities, *Instructional Science*, *44*(3), 197–203, doi:10.1007/s11251-016-9376-y.

Lagun, D., Manzanares, C., Zola, S. M., Buffalo, E. A., & Agichtein, E., D. Lagun, C. Manzanares, S. M. Zola, E. A. Buffalo, and E. Agichtein (2011), Detecting cognitive impairment by eye movement analysis using automatic classification algorithms, *Journal of neuroscience methods*, *201*(1), 196–203, doi: 10.1016/j.jneumeth.2011.06.027.

Lampert, M., and D. L. Ball (1998), *Teaching, Multimedia, and Mathematics: Investigations of Real Practice. The Practitioner Inquiry Series.*, ERIC.

Land, M., N. Mennie, and J. Rusted (1999), The roles of vision and eye movements in the control of activities of daily living, *Perception*, *28*(11), 1311–1328.

Land, M. F., and P. McLeod (2000), From eye movements to actions: how batsmen hit the ball, *Nature neuroscience*, *3*(12), 1340.

Lao, J., S. Miellet, C. Pernet, N. Sokhn, and R. Caldara (2015), imap 4: An open source toolbox for the statistical fixation mapping of eye movement data with linear mixed modeling., *Journal of vision*, *15*(12), 793.

Law, B., M. S. Atkins, A. E. Kirkpatrick, and A. J. Lomax (2004), Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment, in *Proceedings of the 2004 symposium on Eye tracking research & applications*, pp. 41–48, ACM.

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

Leinhardt, G., and J. G. Greeno (1986), The cognitive skill of teaching., *Journal of educational psychology*, *78*(2), 75.

Leonards, U., R. Baddeley, I. D. Gilchrist, T. Troscianko, P. Ledda, and B. Williamson (2007), Mediaeval artists: Masters in directing the observers' gaze, *Current Biology*, *17*(1), R8–R9.

Leshem, S., and Z. Markovits (2013), Mathematics and english, two languages: Teachers' views, *Journal of Education and Learning*, *2*(1), 211.

Lessiter, J., J. Freeman, E. Keogh, and J. Davidoff (2001), A cross-media presence questionnaire: The itc-sense of presence inventory, *Presence: Teleoperators & Virtual Environments*, *10*(3), 282–297.

Levenshtein, V. I. (1966), Binary codes capable of correcting deletions, insertions, and reversals, doi:citeulike-article-id:311174.

Liechty, J., R. Pieters, and M. Wedel (2003), Global and local covert visual attention: Evidence from a Bayesian hidden Markov model, *Psychometrika*, *68*(4), 519–541, doi:10.1007/BF02295608.

Litchfield, D., L. J. Ball, T. Donovan, D. J. Manning, and T. Crawford (2010), Viewing another person's eye movements improves identification of pulmonary nodules in chest x-ray inspection., *Journal of Experimental Psychology: Applied*, *16*(3), 251.

Livingston, C., and H. Borko (1989), Expert-novice differences in teaching: A cognitive analysis and implications for teacher education, *Journal of teacher education*, *40*(4), 36–42.

Mann, D. T., A. M. Williams, P. Ward, and C. M. Janelle (2007), Perceptual-Cognitive Expertise in Sport: A Meta-Analysis, *Journal of Sport and Exercise Psychology*, *29*(4), 457–478, doi:10.1123/jsep.29.4.457.

Mannan, S., K. Ruddock, and D. Wooding (1995), Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images, *Spatial vision*, *9*(3), 363–386.

Mannan, S. K., K. H. Ruddock, and D. S. Wooding (1996), The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images, *Spatial vision*, *10*(3), 165–188.

Marcus, N., B. Cleary, A. Wong, and P. Ayres (2013), Should hand actions be observed when learning hand motor skills from instructional animations?, *Computers in Human Behavior*, *29*(6), 2172–2178.

Martinez-Conde, S., S. L. Macknik, and D. H. Hubel (2004), The role of fixational eye movements in visual perception, *Nature Reviews Neuroscience*, *5*(3), 229–240.

Mason, L., P. Pluchino, and M. C. Tornatora (2015), Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning, *Contemporary Educational Psychology*, *41*, 172–187.

Massy, W. F., and R. Zemsky (1995), *Using information technology to enhance academic productivity*, Educom Washington, DC.

Mathôt, S., F. Cristino, I. D. Gilchrist, and J. Theeuwes (2012), A simple way to estimate similarity between pairs of eye movement sequences, *Jemr*, *5*(1), 1–15, doi:10.16910/jemr.5.1.4.

Mayer, R. E., and C. I. Johnson (2008), Revising the redundancy principle in multimedia learning., *Journal of Educational Psychology*, *100*(2), 380.

McCormick, E. P., C. D. Wickens, R. Banks, and M. Yeh (1998), Frame of reference effects on scientific visualization subtasks, *Human Factors*, *40*(3), 443–451.

McIntyre, N. A., M. T. Mainhard, and R. M. Klassen (2017), Are you looking to teach? Cultural, temporal and dynamic insights into expert teacher gaze, *Learning and Instruction*, *49*, 41–53, doi:10.1016/j.learninstruc.2016.12.005.

Mehlhorn, K., and P. Sanders (2008), *Algorithms and data structures: The basic toolbox*, Springer Science & Business Media.

Memmert, D., D. J. Simons, and T. Grimme (2009), The relationship between visual attention and expertise in sports, *Psychology of Sport and Exercise*, *10*(1), 146–151.

Meyer, A. S., and F. Lethaus (2004), The use of eye tracking in studies of sentence generation, *The interface of language, vision, and action: Eye movements and the visual world*, pp. 191–211.

Milgram, P., and H. W. Colquhoun Jr (1999), A framework for relating head-mounted displays to mixed reality displays, in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 43, pp. 1177–1181, SAGE Publications Sage CA: Los Angeles, CA.

Miller, G. A. (1956), The magical number seven, plus or minus two: Some limits on our capacity for processing information., *Psychological review*, *63*(2), 81.

Müller, S., B. Abernethy, and D. Farrow (2006), How do world-class cricket batsmen anticipate a bowler's intention?, *Quarterly journal of experimental psychology*, *59*(12), 2162–2186.

Nelson, W. W., and G. R. Loftus (1980), The functional visual field during picture viewing., *Journal of Experimental Psychology: Human Learning and Memory*, *6*(4), 391.

Ng, A. Y. (2000), Cs229 lecture notes, *CS229 Lecture notes*, *1*(1), 1–3.

Ng, A. Y., and M. I. Jordan (2002), On Discriminative vs. Generative classifiers: A comparison of logistic regression and naive Bayes., *Advances in neural information processing systems*, pp. 841–848, doi:10.1017/CBO9781107415324.004.

Niebur, E., C. Koch, and C. Rosin (1993), An oscillation-based model for the neuronal basis of attention, *Vision research*, *33*(18), 2789–2802.

Nisbet, S., and E. Warren (2000), Primary school teachers beliefs relating to mathematics, teaching and assessing mathematics and factors that influence these beliefs, *Mathematics Teacher Education and Development*, *2*(34-47).

Nodine, C. F., and H. L. Kundel (1987), Using eye movements to study visual search and to improve tumor detection., *Radiographics*, *7*(6), 1241–1250.

Nodine, C. F., P. J. Locher, and E. A. Krupinski (1993), The role of formal art training on perception and aesthetic judgment of art compositions, *Leonardo*, pp. 219–227.

Noton, D., and L. Stark (1971a), Scanpaths in saccadic eye movements while viewing and recognizing patterns, *Vision Research*, *11*(9), doi:10.1016/0042-6989(71) 90213-6.

Noton, D., and L. Stark (1971b), Scanpaths in Eye Movements during Pattern Perception, *Science*, *171*(3968), 308–311, doi:10.1126/science.171.3968.308.

Olshausen, B. A., C. H. Anderson, and D. C. Van Essen (1993), A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information, *Journal of Neuroscience*, *13*(11), 4700–4719.

Parkhurst, D., K. Law, and E. Niebur (2002), Modeling the role of salience in the allocation of overt visual attention, *Vision Research*, *42*(1), 107–123, doi:10.1016/S0042-6989(01)00250-4.

Parkhurst, D. J., and E. Niebur (2003), Scene content selected by active vision, *Spatial Vision*, *16*(2), 125–154, doi:10.1163/15685680360511645.

Perry, B., D. Tracey, and P. Howard (1999), Head mathematics teachers beliefs about the learning and teaching of mathematics, *Mathematics Education Research Journal*, *11*(1), 39–53.

Peters, R. J., A. Iyer, L. Itti, and C. Koch (2005), Components of bottom-up gaze allocation in natural images, *Vision Research*, *45*(18), 2397–2416, doi:10.1016/j.visres.2005.03.019.

Pietrzyk, M. W., M. F. McEntee, M. E. Evanoff, P. C. Brennan, and C. R. Mello-Thoms (2014), Direction of an initial saccade depends on radiological expertise, in *Medical Imaging 2014: Image Perception, Observer Performance, and Technology Assessment*, vol. 9037, p. 90371A, International Society for Optics and Photonics.

Privitera, C. M. (2006), The Scanpath Theory : its definition and later developments, *Human Vision and Electronic Imaging XI*, *6057*(February 2006), 1–5, doi:10.1117/12.674146.

Privitera, C. M., and L. W. Stark (2000), Algorithms for defining visual regions-of-lnterest: comparison with eye fixations, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(9), 970–982, doi:10.1109/34.877520.

Putnam, R. T. (1987), Structuring and adjusting content for students: A study of live and simulated tutoring of addition, *American educational research journal*, *24*(1), 13–48.

Ramat, A. G., C. Mauri, and P. Molinelli (2013), *Synchrony and diachrony: A dynamic interface*, vol. 133, John Benjamins Publishing.

Rayner, K. (1998), Eye movements in reading and information processing: 20 years of research., *Psychological bulletin*, *124*(3), 372.

Rayner, K., X. Li, C. C. Williams, K. R. Cave, and A. D. Well (2007), Eye movements during information processing tasks: Individual differences and cultural effects, *Vision research*, *47*(21), 2714–2726.

Reingold, E., N. Charness, M. Pomplun, and D. Stampe (2001), Visual Span in Expert chess player: Evidence From Eye Movements, *Psychological Science*, *12*(1), 48, doi:10.1111/1467-9280.00309.

Reingold, E. M., & Sheridan, H. (2011), Eye movements and visual expertise in chess and medicine, in *Oxford handbook on eye movements*, pp. 528–550, Oxford University Press.

Renkl, A. (2014), Toward an instructionally oriented theory of example-based learning, *Cognitive science*, *38*(1), 1–37.

Renninger, L. W., P. Verghese, and J. Coughlan (2007), Where to look next? eye movements reduce local uncertainty, *Journal of Vision*, *7*(3), 6–6.

Richardson, D. C., and M. J. Spivey (2004), Eye-Tracking : Characteristics and Methods Eye-Tracking : Research Areas and Applications Eye-Tracking : Characteristics and Methods, *Biomedical Engineering*, *2*(932839148), 1–32, doi:10.1081/E-EBBE-120013920.

Rieck, K., C. Wressnegger, and A. Bikadorov (2012), Sally: a tool for embedding strings in vector spaces, *Journal of Machine Learning Research*, *13*, 3247–3251.

Robinson, D. A. (1968), The oculomotor control system: A review, *Proceedings of the IEEE*, *56*(6), 1032–1049.

Ross, S. M. (2014), *Introduction to probability models*, Academic press.

Rothkopf, C. A., D. H. Ballard, and M. M. Hayhoe (2007), Task and context determine where you look, *Journal of vision*, *7*(14), 16–16.

Ruiz-Primo, M. A., and E. M. Furtak (2007), Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry, *Journal of research in science teaching*, *44*(1), 57–84.

Saariluoma, P. (1991), Aspects of skilled imagery in blindfold chess, *Acta psychologica*, *77*(1), 65–89.

Sadowski, W., and K. Stanney (2002), Presence in virtual environments., in *Human factors and ergonomics. Handbook of virtual environments: Design, implementation, and applications*, pp. 791–806, Lawrence Erlbaum Associates Publishers.

Salton, G., A. Wong, and C.-S. Yang (1975), A vector space model for automatic indexing, *Communications of the ACM*, *18*(11), 613–620.

Salzman, M. C., C. Dede, R. B. Loftin, and J. Chen (1999), A model for understanding how virtual reality aids complex conceptual learning, *Presence: Teleoperators & Virtual Environments*, *8*(3), 293–316.

Savelsbergh, G. J., A. M. Williams, J. V. D. Kamp, and P. Ward (2002), Visual search, anticipation and expertise in soccer goalkeepers, *Journal of sports sciences*, *20*(3), 279–287.

Schmittmann, V. D., C. V. Dolan, H. L. van der Maas, and M. C. Neale (2005), Discrete latent markov models for normally distributed response data, *Multivariate Behavioral Research*, *40*(4), 461–488.

Schriver, A. T., D. G. Morrow, C. D. Wickens, and D. A. Talleur (2008), Expertise differences in attentional strategies related to pilot decision making, *Human Factors*, *50*(6), 864–878.

Schutz, A. C., D. I. Braun, and K. R. Gegenfurtner (2011), Eye movements and perception: A selective review, *Journal of Vision*, *11*(5), 9–9, doi:10.1167/11.5.9.

Seidel, T., and K. Sturmer (2014), Modeling and Measuring the Structure of Professional Vision in Preservice Teachers, *American Educational Research Journal*, *51*(4), 739–771, doi:10.3102/0002831214531321.

Silverman, B. W. (2018), *Density estimation for statistics and data analysis*, Routledge.

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

Spivey, M. J., and R. Dale (2011), Eye movements both reveal and influence problem solving, in *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, chap. 30, pp. 551–562, Oxford University Press.

Star, J. R., and S. K. Strickland (2008), Learning to observe: Using video to improve preservice mathematics teachers ability to notice, *Journal of mathematics teacher education*, *11*(2), 107–125.

Stockero, S. L., R. L. Rupnow, and A. E. Pascoe (2017), Learning to notice important student mathematical thinking in complex classroom interactions, *Teaching and Teacher Education*, *63*, 384–395, doi:10.1016/j.tate.2017.01.006.

Stokking, K., F. Leenders, J. De Jong, and J. Van Tartwijk (2003), From student to teacher: Reducing practice shock and early dropout in the teaching profession, *European journal of teacher education*, *26*(3), 329–350.

Stürmer, K., T. Seidel, K. Müller, J. Häusler, and K. S. Cortina (2017), What is in the eye of preservice teachers while instructing? An eye-tracking study about attention processes in different teaching situations, *Zeitschrift für Erziehungswissenschaft*, *20*(S1), 75–92, doi:10.1007/s11618-017-0731-9.

Tanenhaus, M. K., C. Chambers, J. E. Hanna, and M. Hall (2004), Referential domains in spoken language comprehension: Using eye movements to bridge the product and action traditions, *The interface of language, vision, and action: Eye movements and the visual world*, pp. 279–317.

Tatler, B. W., and B. T. Vincent (2009a), The prominence of behavioural biases in eye guidance, *Visual Cognition*, *17*(6-7: Eye Guidance in Natural Scenes), 1029–1054.

Tatler, B. W., and B. T. Vincent (2009b), The prominence of behavioural biases in eye guidance, *Visual Cognition*, *17*(6-7), 1029–1054.

Tatler, B. W., R. J. Baddeley, and I. D. Gilchrist (2005), Visual correlates of fixation selection: Effects of scale and time, *Vision Research*, *45*(5), 643–659, doi:10.1016/j.visres.2004.09.017.

Tatler, B. W., N. J. Wade, H. Kwan, J. M. Findlay, and B. M. Velichkovsky (2010), Yarbus, eye movements, and vision, *i-Perception*, *1*(1), 7–27.

Tatler, B. W., M. M. Hayhoe, M. F. Land, and D. H. Ballard (2011), Eye guidance in natural vision: Reinterpreting salience, *Journal of Vision*, *11*(5), 5–5, doi:10.1167/11.5.5.

Theeuwes, J. (1994), Stimulus-driven capture and attentional set: selective search for color and visual abrupt onsets., *Journal of Experimental Psychology: Human perception and performance*, *20*(4), 799.

Thomas, L. C., and C. D. Wickens (2001), Visual Displays and Cognitive Tunneling: Frames of Reference Effects on Spatial Judgments and Change Detection, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *45*(4), 336–340, doi:10.1177/154193120104500415.

Thompson, A. G. (1992), Teachers' beliefs and conceptions: A synthesis of the research., in *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics*, pp. 127–146, Macmillan Publishing Co, Inc.

Treisman, A. (1982), Perceptual grouping and attention in visual search for features and for objects., *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 194.

Treisman, A. (1983), The role of attention in object perception, in *Physical and biological processing of images*, pp. 316–325, Springer.

Treisman, A. M., and G. Gelade (1980), A feature-integration theory of attention, *Cognitive psychology*, *12*(1), 97–136.

Tseng, P. H., I. G. Cameron, G. Pari, J. N. Reynolds, D. P. Munoz, and L. Itti (2013), High-throughput classification of clinical populations from natural viewing eye movements, *Journal of Neurology*, *260*(1), 275–284, doi:10.1007/s00415-012-6631-2.

Tsotsos, J. K., S. M. Culhane, W. Y. K. Wai, Y. Lai, N. Davis, and F. Nuflo (1995), Modeling visual attention via selective tuning, *Artificial intelligence*, *78*(1-2), 507–545.

Ulusoy, I., and C. M. Bishop (2005), Generative versus discriminative methods for object recognition, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, *2*, 258–265, doi:10.1109/CVPR.2005.167.

Van den Bogert, N., J. van Bruggen, D. Kostons, and W. Jochems (2014), First steps into understanding teachers' visual perception of classroom events, *Teaching and Teacher Education*, *37*, 208–216, doi:10.1016/j.tate.2013.09.001.

Van Es, E. A., and M. G. Sherin (2002), Learning to notice: Scaffolding new teachers interpretations of classroom interactions, *Journal of Technology and Teacher Education*, *10*(4), 571–596.

van Gog, T., and N. Rummel (2010), Example-based learning: Integrating cognitive and social-cognitive research perspectives, *Educational Psychology Review*, *22*(2), 155–174, doi:10.1007/s10648-010-9134-7.

Van Gog, T., and N. Rummel (2010), Example-based learning: Integrating cognitive and social-cognitive research perspectives, *Educational Psychology Review*, *22*(2), 155–174.

Van Gog, T., F. Paas, J. J. Van Merriënboer, and P. Witte (2005), Uncovering the problem-solving process: Cued retrospective reporting versus concurrent and retrospective reporting., *Journal of Experimental Psychology: Applied*, *11*(4), 237.

van Gog, T., H. Jarodzka, K. Scheiter, P. Gerjets, and F. Paas (2009), Attention guidance during example study via the model's eye movements, *Computers in Human Behavior*, *25*(3), 785–791, doi:10.1016/j.chb.2009.02.007.

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

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, doi:10.1016/j.jmp.2011.08.002.

Visser, I., M. E. Raijmakers, and P. Molenaar (2002), Fitting hidden markov models to psychological data, *Scientific Programming*, *10*(3), 185–199.

Viviani, P. (1990), Eye movements in visual search: Cognitive, perceptual, and motor control aspects, *Eye movements and their role in visual and cognitive processes*, pp. 353–383.

Vogt, S., and S. Magnussen (2007), Expertise in pictorial perception: eye-movement patterns and visual memory in artists and laymen, *Perception*, *36*(1), 91–100.

Wade, N. J., and B. W. Tatler (2011), Origins and applications of eye movement research, in *The Oxford Handbook of Eye Movements*, edited by S. P. Liversedge, I. D. Gilchrist, and S. Everling, chap. 2, pp. 17–43, Oxford University Press.

Walls, G. (1962), The evolutionary history of eye movements, *Vision Research*, *2*(1-4), 69–80.

Waters, A. J., G. Underwood, and J. M. Findlay (1997), Studying expertise in music reading: Use of a pattern-matching paradigm, *Perception & psychophysics*, *59*(4), 477–488.

Williams, A. M., P. Ward, J. M. Knowles, and N. J. Smeeton (2002), Anticipation skill in a real-world task: measurement, training, and transfer in tennis., *Journal of Experimental Psychology: Applied*, *8*(4), 259.

Wolff, C. E., H. Jarodzka, and H. P. Boshuizen (2017), See and tell: Differences between expert and novice teachers' interpretations of problematic classroom management events, *Teaching and Teacher Education*, *66*, 295–308, doi: 10.1016/j.tate.2017.04.015.

Yamamoto, T., and K. Imai-Matsumura (2013), Teachers' Gaze and Awareness of Students' Behavior: Using An Eye Tracker, *Comprehensive Psychology*, *2*, 01.IT.2.6, doi:10.2466/01.IT.2.6.

Yantis, S., and H. E. Egeth (1999), On the distinction between visual salience and stimulus-driven attentional capture., *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 661.

Yantis, S., and J. Jonides (1984), Abrupt visual onsets and selective attention: evidence from visual search., *Journal of Experimental Psychology: Human perception and performance*, *10*(5), 601.

Yantis, S., and J. Jonides (1996), Attentional capture by abrupt onsets: new perceptual objects or visual masking?, *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1505–1513.

Yarbus, A. L. (1967a), Eye movements during perception of complex objects, in *Eye Movements and Vision*, pp. 389–390, Springer New York, doi:10.1016/0028-3932(68)90012-2.

Yarbus, A. L. (1967b), *Eye movements and vision*, Plenum Press New York.

Ylitalo, A.-k. (2017), Statistical Inference for Eye Movement Sequences using Spatial and Spatio-temporal Point Processes, Ph.D. thesis, University of Jyv askyl a.

Zakaria, E., and N. Musiran (2010), Beliefs about the nature of mathematics, mathematics teaching and learning among trainee teachers, *Social Sciences*, *5*(4), 346–351.

Zangemeister, W., K. Sherman, and L. Stark (1995), Evidence for a global scanpath strategy in viewing abstract compared with realistic images, *Neuropsychologia*, *33*(8), 1009–1025.