

Chapter 13

Methodology and Application of Eye-Tracking Techniques in Science Education

Miao-Hsuan Yen and Fang-Ying Yang

Abstract Eye-tracking techniques have the potential to reveal learning processes and problem solving strategies when dealing with science content. For instance, locations, durations, and numbers of gazes in predefined areas of interest together with transitions between areas indicate the degree of attention allocation to these areas. These measures combined with achievement tests, thinking aloud, or interviews can shed light on how or why learners gain or do not gain knowledge. This chapter aims to introduce issues in science education that can be explored with eye-tracking techniques and methodological concerns that should be considered when conducting research and interpreting data. The first section of this chapter outlines the methodological issues and the second section presents a survey of empirical studies conducted in Taiwan. Specifically, since material with science content usually consists of multiple representations, attentional distribution in text and illustration is one of the research foci. Furthermore, how participants' prior knowledge and additional cues in the material guide attention allocation is also investigated. Implications for future development are discussed in the last section.

Investigating online cognitive processes during learning and problem solving is one of the key issues in science education. Thinking aloud protocols (Ericsson and Simon 1993), videotaping, and computer logs have been utilized to reveal the online process. Recently, eye-tracking (van Gog and Scheiter 2010) and neuroimaging Liu and Huang (in press) techniques have been applied in educational studies. These techniques can record the natural behavior involved in performing the task without additional demands (such as concurrent verbal reports). For instance, the eye-tracking data provide valuable information about *which* parts of the material are attended to and *how long* they are processed.

The purpose of this chapter is to document the applications of eye-tracking techniques to study issues in science education with a focus on research conducted

M.-H. Yen (✉) · F.-Y. Yang
Graduate Institute of Science Education, National Taiwan Normal University, Taipei, Taiwan
email: myen@ntnu.edu.tw

© Springer Science+Business Media Singapore 2016
M.-H. Chiu (ed.), *Science Education Research and Practices in Taiwan*,
DOI 10.1007/978-981-287-472-6_13

249

in Taiwan. Since the first study in 2009, 15 journal articles on empirical studies executed in Taiwan have been published to date. The accelerated and flourishing development in educational research applying eye tracking in Taiwan is partly due to the support of multidisciplinary research projects from the National Science Council since 2007. Furthermore, researchers were intrigued by understanding the process of learning through online measurement rather than solely observing the end result of the intervention. The survey presented in the second section of this chapter illustrates what can be investigated with eye-tracking techniques, and reveals potential future exploration. Since learning material and problems with science content may consist of multiple representations so that information selection and integration may be demanded, multiple factors should be considered to interpret the complicated cognitive processes revealed by eye-tracking data. Therefore, several methodological issues of eye-tracking techniques that should be considered are briefly introduced in the first section. However, this introduction is by no means comprehensive, and is limited in scope. To gain more information and insights, readers can refer to other books (such as Duchowski 2007; Holmqvist et al. 2011) and review articles (such as Rayner 1998, 2009).

13.1 Methodological Concerns: The Cognitive, Physiological, and Technical Bases of Eye Tracking

Since the proposal of the eye–mind assumption by Just and Carpenter (1980), eye-tracking techniques have been adopted by researchers in various fields to reveal the underlying processes during visual tasks. The eye–mind assumption suggests that the eyes remain fixated on the region being processed. Thus, gaze duration in that area can be used as an index of processing time. However, parafoveal processing was found in subsequent studies (see Rayner 1998 for a review). In other words, stimuli that are not directly fixated on (i.e., projected to the fovea) can also be processed. For example, as Just and Carpenter acknowledged, stimuli from previous fixation(s) may still be processed and integrated with the information obtained from the present fixation. In addition, nearby stimuli that have not been fixated on can be processed to some extent (which is usually referred to as parafoveal preview). The region from which useful information can be extracted during fixations is called the perceptual span, the size of which varies with task demands, task difficulty, and perceivers' experience and skill. Many researchers in the fields of reading, scene perception, and visual search have investigated what information can be extracted from the perceptual span (especially the parafovea) and what influences the decision of the next saccadic targets. Although the mechanism of eye movements during visual tasks is more complicated than that suggested by the eye–mind assumption, it is still a useful approximation because, for example, gaze durations are influenced by linguistic factors such as the frequency and predictability of the fixated word (Rayner et al. 2003), indicating immediate cognitive (linguistic) processing during fixation.

To sum up, gaze duration may reflect the time for processing stimuli in the fixated and nearby areas, as well as the information obtained from the previous fixations. That is, it can be an assemblage of encoding new information and integrating with recently extracted information and prior knowledge.

13.1.1 Eye-Tracker System

There are several eye-tracking techniques, such as electrooculography (EOG), search coils (scleral contact lens), and video-based eye trackers (Duchowski 2007). Among these techniques, video-based eye trackers are often used in the fields of cognitive psychology, education, sports, advertising, consumer behavior, etc. The movement of the pupil, corneal reflection, and some facial features are tracked by video-based eye trackers. The system usually consists of a host PC (where eye movements are recorded and monitored) and a display PC (where experimental stimuli are presented and screen capture may be conducted), which are usually connected by an Ethernet cable. Video-based eye trackers can be desktop and head mounted. Desktop-mounted eye trackers can track the eyes remotely; nevertheless, a chin rest is used for some trackers. Some head-mounted eye trackers can be used when participants freely move in the surroundings; eye movements are recorded by a connected device and might be monitored in another computer through a wireless connection. Sometimes, a scene camera is set up to record what the participants are looking at, which can also be recorded by screen capture software if the material is presented by computer. Eye trackers also differ in their sampling rates (e.g., 30, 60, 120, 250, 500, 1000 Hz) and algorithms to parse sampled data into fixations and saccades (or non-fixations), which will be described in the following subsections. After setting up the tracker, calibration is usually conducted to obtain the point of regard on the monitor or in the real scene. Typically, several fixation points are presented one-by-one in the center and corners of the monitor or the scene to be viewed. The order of the location of the fixation points is usually random. The participant is instructed to fixate on those fixation points in turn. A validation procedure is then conducted to verify the accuracy. Some oculomotor errors (e.g., $0.4\sim 1^\circ$) are acceptable, but the criterion varies among laboratories and can be set according to the specific task demands and the grain size of the areas of interest. For mobile eye trackers, calibration can also be conducted by recording some markers in the real scene.

13.1.2 Basic Patterns of Eye Movements

Due to the decrease in visual acuity (i.e., the drop in the density of photoreceptors in the retina) from the fovea to the periphery, people need to move their eyes to sample information for further processing (Findlay and Gilchrist 2003). The eye

movement pattern during visual tasks is characterized by a series of alternating fixations and saccades. The eyes move rapidly during saccades (the duration and maximum velocity of which depend on saccade length, which is called the saccadic main sequence) while they remain *relatively* static during fixation. There are fixational eye movements to prevent photoreceptors from saturation (Martinez-Conde and Macknik 2007). It follows that there is no clear-cut distinction between fixations and saccades (Inhoff and Radach Inhoff 1998), although several criteria can be set to distinguish them. Some trackers segment raw data based on detecting saccades, whereas others based on detecting fixations. The raw data provided by video-based eye trackers are x- and y-positions as well as the diameter of the pupil captured in each sample. Some eye trackers adopt a velocity threshold. That is, if the speed and acceleration of the eye movements are higher than the predefined thresholds, these samples are categorized as saccades whereas others are treated as fixations. A typical saccade duration for a reading is 20–50 ms. If the sampling rate of the eye tracker is not high enough, it may be difficult to detect saccades. For example, with a 60-Hz sampling rate, the interval between two consecutive samples is 16.7 ms. Then, it cannot detect a 20-ms saccade. In some eye trackers or software, duration and distance threshold are used to aggregate samples into fixations. For instance, when the distance between consecutive samples is less than 0.5° of visual angle for 80 ms, these samples are treated as fixations, whereas others are treated as non-fixations. In some software, for instance, the start of a fixation is defined when six consecutive samples are located within 0.5° , and the end of a fixation is defined when three consecutive samples are located more than 1° apart. The types of threshold are usually determined by the particular eye trackers. Nevertheless, researchers can design their own software to process the raw data.

Information is mainly extracted during fixations, although it is continuously processed during saccades and blinks (Inhoff and Radach 1998; Rayner 1998). Therefore, fixation is the main concern in most cognitive studies. There is large variation in how long the eyes remain fixated. The range of fixation duration can be from shorter than 50 ms to longer than 500 ms (Rayner 1998). Although fixation duration can be influenced by cognitive processing, very short or long fixations may be caused by factors unrelated to the current cognitive task. As it takes time for visual information to travel from the eyes to the brain (about 60 ms, Reichle and Reingold 2013), short fixations may result from oculomotor errors and may not be enough for visual processing. Thus, they are combined with consecutive fixations or discarded in some studies. On the other hand, very long fixations are sometimes discarded because the participants may be absentminded or distracted by something unrelated to the task. The criteria to exclude duration data from analysis vary among cognitive tasks. The typical mean fixation duration in different tasks can be taken into consideration. For example, the mean fixation durations during silent reading, reading aloud, and scene perception are 225–250, 275–325, and 260–330 ms, respectively (Rayner 2009). Whether or not additional criteria for fixation durations are used, and the particular cut-off values are determined by the research focus and task used.

13.1.3 Data Processing

There are a variety of eye movement measures. In general, these measures concern the spatial, temporal, frequency, and sequential dimensions, namely, how far the eyes move and which part of the stimulus is fixated on, how long and how many times it is fixated on, and the pattern of transitions among different parts of the material (Holmqvist et al. 2011; Lai et al. 2013; Radach and Kenney 2004). Sometimes, new measures are created to reflect the cognitive process specific to the task. Although analyzing all fixations in the unsegregated material can be meaningful, most of the time, investigating eye movement patterns in and between specific target areas in the material can further reveal the attentional distribution and strategies used during the task. These target areas are called the areas of interest (AOIs). In some software, AOIs are also called regions of interest (ROIs), interest areas (IAs), and look zones (LZs), to name a few. In basic reading research, words are used as AOIs. However, in studies of high-level processing, clauses and sentences are treated as AOIs. Larger segments, such as paragraphs and diagrams, can be used as AOIs in applied studies. Diagrams can be further segmented into the different elements that are depicted. When different AOIs in a study vary in size, their areas have to be taken into consideration when the duration and frequency of the fixations are compared among AOIs. For example, the number of fixations and fixation duration per pixel can be calculated. In addition, transitions between AOIs can also be investigated. Whether or not number of transitions between AOIs A and B is significantly more than that between A and C can be an indicator of the strategies used. The techniques to calculate transitions between AOIs are described comprehensively by Holmqvist et al. (2011).

Once AOIs are defined, the number and duration of fixations in each AOI can be calculated and then comparisons between AOIs can be made. At one extreme, an AOI can be skipped (i.e., not fixated on at all) when the stimuli are presented, presumably because it can be easily processed parafoveally or it does not attract the participant's attention; the corresponding measure is fixational probability or skipping rate. If it is fixated on, besides calculating the total number and duration of fixations, these fixations can be distinguished into first pass, second pass, etc., to reflect different stages of processing (Inhoff and Radach 1998; Radach and Kennedy 2004). First-pass fixations are the group of fixations in which the AOI is inspected for the first time after stimulus onset; there might be multiple fixations before leaving the AOI. If the eyes return to the AOI after leaving it, second-pass fixations can be identified. The AOI can further receive third- and fourth-pass fixations, and so on. Parenthetically, passes are named runs or visits in some software and studies. First-pass measures seem to reflect the initial encoding of the stimuli in the AOI. On the other hand, if the AOI receives additional pass(es), it may reflect later integration with information obtained from other AOIs. For instance, Kaakinen et al. (2003) found that readers with high working memory capacity (WMC) spent more time processing relevant than irrelevant information during the first pass, whereas those with low WMC spent less time processing irrelevant information during the

first pass but increased look-back time (second pass) on relevant information. Distinguishing the first pass from additional passes reveals the detailed strategies used by different readers.

The nomenclature of eye movement measures varies across software and laboratories. In the field of basic reading research where word recognition during reading is one of the research foci, fine-grained measures for first-pass fixations are specified. For instance, the sum of the duration of all first-pass fixations is called (first-pass) gaze duration (GD). In some software, GD may be called first-run dwell time. The duration of the very first fixation during the first pass is called first fixation duration (FFD). Other first-pass fixations are called refixations. Fixations other than first-pass ones can be labeled as regression, look backs, rereading, or revisits. For instance, look-back time, rereading time/rate, and regression probability can all be calculated. When all fixations are considered, the number of fixations and the number of gazes/passes/runs can be measured. In addition, the sum of durations of all fixations is called total reading time, total viewing time, total fixation duration, total inspection time, dwell time, etc.

Besides AOIs, periods of interest can be investigated as well. For example, what the learners explore during the initial 30 s may indicate the areas that most attract their attention. Similarly, the number of fixations or time before a certain AOI is fixated on after stimuli onset can be calculated to reveal the search efficiency. This duration of time is called latency or time to first fixation. In addition, researchers may be interested in where participants fixated before they made decisions. It should be noted that if presentation duration of the material is determined by the participants, total time in trial has to be considered when comparing fixation durations among participants. For instance, a 1-s total viewing time in an AOI can be long for a 5-s presentation, although it can be short for a 15-s presentation.

In the articles mentioned in the second section, several eye movement measures as well as transition patterns were reported. Two of the measures were total fixation duration and total fixation count. In some studies, the sizes of different AOIs were taken into consideration. In addition, some studies reported fixation counts during reinspection and fixation duration during the first pass. The other two frequently reported measures were mean fixation duration and mean saccade length. The design and findings of these studies are described in the second section and summarized in the Appendix.

13.1.4 Factors that May Influence Eye Movement Patterns

Eye movement is a natural behavior in tasks that involve visual processing. Factors along the information processing pathways might influence *where* the eyes locate and *how long* they remain fixated. As the review by Henderson (2003) stated, there is bottom-up information embedded in the scene and top-down information generated by the cognitive system. A uniform region with single color, brightness, texture, etc., may be uninformative, whereas a sharp contrast in these dimensions may

signify a boundary between multiple objects or their surroundings. Thus, saliency can play an important role in parsing the scene and signifying potential targets in it. On the other hand, task demands such as searching for a specific target among distractors, viewing the scene to make a decision (e.g., to evaluate whether or not to buy the product or how effective the advertisement is), or comprehending a learning material can influence attention distribution in inspecting the stimulus. Knowledge of the target or the scene (such as a living room or a hospital) can indicate the cognitive relevance of the stimuli and locations in the scene, and guide the allocation of attention to them. For example, world knowledge suggests that a watch is more likely to be located on a table, while it is nearly impossible that it will be floating in the air or hung from the ceiling. In addition, task-relevant knowledge suggests how to process the scene. For instance, when driving a car, frequently looking through the windscreen or mirrors is needed to check the traffic lights or the behavior of other vehicles or pedestrians. To sum up, stimulus characteristics and knowledge jointly determine how the presented material or the scene is processed and how the eyes extract information for further processing.

Several descriptive or computational models have been constructed to illustrate the coordination of visual processing, cognitive processing, and oculomotor systems during a variety of tasks. In models concerning basic reading process, it is generally agreed that linguistic processing plays an important role in the *where* and *when* decisions of eye movements (see Reichle et al. 2003, for a review). It should be noted that as it takes time to initiate and execute a saccade (with latency around 150 ms) once the saccadic target is decided, linguistic processing and saccade programming overlap in time during fixation. However, the period in which linguistic processing can influence whether the same target is refixated on or not is quite limited (e.g., in a typical 250-ms fixation, the *where* decision may be made during the initial 100 ms). Thus, parafoveal processing may be important during reading. As expected, models propose different mechanisms of how linguistic processing (foveal and parafoveal) is conducted and then influences oculomotor control, which is beyond the scope of this chapter. However, these models mainly concern word recognition (and factors such as word frequency and predictability from context) in the linguistic processing, with recent attempts to take higher-order processing (such as sentence-level comprehension and reading goals) into consideration (Radach et al. 2008; Reichle et al. 2009).

Processing educational material can be complicated. For example, multiple representations usually coexist; whether “a picture paints a thousand words” or not may largely depend on task demands and learners’ prior knowledge. In addition, learners may not have enough knowledge to determine relevant parts of the material so that their information extraction may be influenced by perceptual saliency, information densities, etc. Although cognitive processing is the main concern in educational research, perceptual-level and stimulus-based properties should be taken into consideration. Eye movement patterns can be influenced by both perceptual and cognitive levels of processing. Furthermore, although fixation duration may reflect how much attention is allocated to the fixated and nearby areas it is fixated on and attended to probably because it is interesting, relevant, or difficult, which leads to

different interpretations. Combining eye movement measures with others such as thinking aloud, retrospective interviews, or comprehension tests may resolve the ambiguity. Moreover, considering a set of eye movement measures together instead of interpreting a single measure can be useful. For example, a long mean fixation duration together with a long mean saccade length may indicate that information is extracted from a large surrounding area during each fixation (Rayner et al. 2001). On the contrary, a long mean fixation duration with a short mean saccade length may imply that the material has high information density or is difficult to process (Chen et al. 2014). In the following subsections, some models and studies on how saliency and task demands influence scene perception are introduced.

13.1.4.1 Perceptual Saliency Versus Cognitive Relevance

Concerning scene perception and visual search, models disagree on whether saliency or higher order cognitive processing (such as task demands) dominate in where and when decisions (Tatler 2009). As Tatler's introduction to the special issue in *Visual Cognition* mentioned, models arguing against the pure saliency model have been developed recently. Factors that may influence fixation location include prior knowledge of similar scenes and targets (e.g., where a certain target is more likely located in a particular scene), oculomotor tendencies (e.g., central fixation bias, Tatler 2007), and perceptual saliency. Two of the models that incorporate higher level factors are briefly introduced as follows.

Navalpakkam and Itti (2005) constructed a computational model in which saliency and task-relevance maps are combined into an attention guidance map from which the next attentional (saccade) target can be selected through a winner-takes-all competition. First of all, task specification (e.g., to find what the man is eating) activates relevant prior knowledge in the long-term memory (e.g., knowledge about food, humans, and eating behavior) and then biases associated features (i.e., a certain color or shape) in the perceived scene in the visual system. Second, a saliency map is computed with weighting for associated features. At the same time, a task-relevance map is also computed taking the likely or relevant locations (e.g., the likely locations of a man and the food he is eating) into consideration. Third, a potential object or location is attended to and recognized, and then its relevance to the task is evaluated. Finally, working memory is updated. If the attended region is not relevant, then look for the next relevant region. If it is somewhat relevant (e.g., a man is found), then look for a more relevant region (e.g., his mouth or hand). The cycle will iterate until the task is completed.

On the other hand, Henderson et al. (2009) proposed a cognitive relevance theory in which salience helps to build an unranked scene representation upon which potential saccade targets are weighted by cognitive relevance. In other words, saliency simply makes it easy to parse the scene into different objects or regions; it is the cognitive relevance that determines the fixation location. In their study, participants were instructed to search for a non-salient target object in a real scene with other salient objects or regions. Nevertheless, participants rarely looked at the salient region

(less than 12% of time). In addition, chance is high that participants fixate on the target regions immediately after the stimulus onset (more than 87% of the time).

With a different task, Kaakinen et al. (2011) demonstrated both the effects of perceptual saliency and cognitive relevance on scene viewing. In their study, undergraduate students viewed sets of photos about the exterior and interior of three houses with one of the three types of instruction. The homebuyer group was instructed to decide which house they would like to buy, the burglar group was instructed to decide which house to break into, whereas the memory group was instructed to prepare for a memory test about these pictures. The eye movement patterns of the memory group served as the baseline because the group viewed the photos without any particular perspective. An additional group of participants selected regions in each photo that were relevant to either homebuyers or burglars. Saliency was analyzed by software. For each perspective (homebuyers or burglars), there were salient-relevant, non-salient-relevant, salient-irrelevant, and non-salient-irrelevant regions in each photo. The results showed that, compared to the baseline, viewing perspective lengthened the total viewing time and first-pass dwell time, as well as raised the number of visits in perspective-relevant areas, whereas it shortened the total viewing time and lowered the number of visits in perspective-irrelevant areas. However, it did not decrease the number of visits in salient-irrelevant areas. For the memory group, saliency was the only factor to increase total viewing time, number of visits, and first-pass dwell time. For the other groups with specific viewing perspectives, saliency and relevance had general additive and sometimes interactive effects (which may result from differences in burglar-relevant and homebuyer-relevant regions). In general, total viewing time was 1047, 844, 620, and 324 ms, and number of visits was 1.77, 1.54, 1.45, and 0.83 in the salient-relevant, non-salient-relevant, salient-irrelevant, and non-salient-irrelevant areas, respectively. The pattern of results in the first-pass dwell time for homebuyers and burglars was less obvious; the effect of perspective relevance was observed whereas the effect of saliency was not. Furthermore, probabilities of fixating relevant or salient areas during the first ten fixations were analyzed. For the memory group, only the saliency effect was detected. On the other hand, for the groups with specific perspectives, there was an interaction between relevance and fixation order. For the first fixation, only a main effect of saliency was found. From the second fixation on, both effects of saliency and relevance were observed. To summarize, although saliency influenced the location of the first fixation, both saliency and relevance determined subsequent fixation location, number of visits, and total viewing time. That is, in a task without prespecified targets (such as search targets in the study of Henderson et al. 2009), an immediate effect of saliency followed by a rapid relevance effect and their joint effects were observed.

13.1.4.2 Influence of Task Demands

The influence of task demands was demonstrated by the two studies (Henderson et al. 2009; Kaakinen et al. 2011) exemplified in 13.1.4.1 Though relevance might

overwrite the effect of saliency when a target was specified before the scene was viewed (Henderson et al. 2009), they might on the other hand cooperate in a task without specific targets (Kaakinen et al. 2011). That eye movement behavior is influenced by task demands can be further demonstrated in two studies conducted by Rayner et al. (2001, 2008). In their studies, real advertisements taken from magazines were used as the material, eight of which were about cars and eight of which were about skin-care products. In their first study, adult participants were instructed to pretend that they had just moved to another country and needed to buy either a new car or some skin-care products (half of the participants were assigned to one of the groups). In their second study, participants were instructed to rate how they liked each of the advertisements or judged how effective these advertisements were. In this second study, an additional 32 advertisements were used together with the 16 advertisements used in the first study. In the advertisements, there were pictures (some were photos of the products whereas others were related to the products) and text (with large and small print). In the first study, in which the participants pretended to be buyers for cars or skin-care products, they had longer looking time and more fixations on the corresponding relevant advertisements. They also had longer total looking time and more fixations on the text regions than the pictures even when the size (pixels) of the different regions was taken into consideration. In addition, the mean fixation durations and saccade lengths were longer in the pictures than in the text regions, suggesting that more information could be extracted from the pictures than the text regions in each fixation. Participants tended to read the large print, followed by the small print and then the pictures, although they did occasionally fixate on the pictures before reading the small print. When they read the text, they did not read all of it if the text was long, suggesting that they did not carefully read the text to fully comprehend it as they might do in a typical reading task.

On the contrary, in the second study, Rayner et al. (2008) found quite different patterns of results. First of all, when participants were asked to rate how much they liked the advertisement or to evaluate the effectiveness of the advertisements, they spent more time fixating and made more fixations on the pictures (61%) rather than on the text areas (39%). Separately analyzing the 16 advertisements for cars and skin-care products and the others for different products revealed that both task goals and specific material influenced eye movement patterns. The participants spent half of their time on the text and picture areas when they viewed the original 16 advertisements, but spent more time on the picture than on the text areas for the other material. Rayner et al. (2008) suggested that presumably because the advertisements about cars and skin-care products had a larger number of words, participants spent more time on the text for these advertisements than other advertisements used in the second study. Second, the probability of fixating on the picture in the first fixation after the stimulus onset was about 69%, whereas that on the text areas was about 16%. Nevertheless, consistent with their previous study (Rayner et al. 2001), the participants had longer mean fixation durations and saccade lengths for the pictures than for the text areas. In addition, no frequent alternation between the text and picture areas was observed. A comparison of the two studies revealed that the participants distributed their attention distinctively according to different task

instructions. They also varied their sequence to extract information; specifically, they mainly relied on information carried in the text and then sought confirmation from the pictures in one study or vice versa in another.

The findings of Rayner et al. (2001, 2008) suggest that when viewing the same set of material, the attentional distribution and order of allocation among multiple representations change corresponding to task demands. To complicate the issue, when comprehending educational material, the “targets” may change over time. For example, in the material describing how a natural phenomenon takes place or how a machine works, there is a sequence of events happening during the process. It follows that the eyes and the mind change their foci among different parts of the material and then integrate new information with that obtained in previous fixations from different areas. The application of eye-tracking techniques in science education is in its infancy compared to the number of studies in basic reading, visual search, and scene perception, although some related work has been reported in the literature (e.g., Hegarty and Just 1993). At the exploratory stage, basic eye movement patterns are first identified, with increasing complexity being gradually considered in more recent studies. In the next section, studies related to science education conducted in Taiwan are documented.

13.2 Application of Eye-Tracking Techniques in Science Education in Taiwan

In this section, studies concerning science education and using eye-tracking techniques are discussed with an emphasis on research conducted in Taiwan. A total of 15 articles were selected by first identifying active researchers in the field and then searching their publication lists for relevant articles. The researchers were identified from the review article by Lai et al. (2013), abstract books of recent European Conferences on Eye Movements, and a special issue of the *International Journal of Science and Mathematics Education* (2014). In addition, research interests and publication lists of faculty members in departments and institutes of (science) education in universities in Taiwan were searched to identify active researchers. Empirical articles published in local and international peer-reviewed journals were then included in this survey.

Four major issues were investigated in these 15 articles. In most studies, the material consisted of verbal information and visualizations. Specifically, seven of them explored how participants comprehended or learned from material with written text and illustrations. Two studies compared the effects of presentation modes (e.g., animation versus simulation, narration versus text) on learning outcomes and eye movement behavior. One study documented how students inspected PowerPoint slides in a real classroom. Five studies demonstrated how participants solved problems with science content. In their data analysis, attentional distribution (manifested by fixation duration and frequency) between the illustration and text areas was one of the research foci. Another focus was the transition pattern among AOIs, revealing

the strategies the participants used. In addition, some of the studies investigated how participants' prior knowledge influenced the way they processed the material. Also, the effects of coloring relevant parts or adding arrows on reading comprehension were investigated by Wu and colleagues (Chen and Wu 2012; Jian and Wu 2012; Jian et al. 2014). These 15 articles are outlined and discussed in the following subsections with their experimental manipulation, types of material, characteristics of participants, and findings summarized in the Appendix.

13.2.1 Learning from Material with Text and Illustrations

Liu and Chuang (2011) demonstrated that when nonscience-major undergraduate students read text about the formation of winds and atmospheric pressure systems with illustrations for comprehension, they had a greater total number of fixations and longer total fixation durations in the text area than in the illustration area. This pattern remained when the sizes of different areas were taken into consideration (i.e., calculating fixation and duration densities). In addition, a similar pattern of results was found even when the size of the text area was smaller than that of the illustration area (because one of the concepts was presented step-by-step over five pages, the number of words was reduced on each page). Concerning transitions between text and illustration areas, participants inspected the illustration when they read the text describing the kinematics of the system (e.g., clockwise movement of the air). A global transition pattern was found in one concept (atmospheric pressure), in which participants fixated on the illustration more often following reading the end of a sentence than after reading any other parts of the sentences. Thus, readers seemed to build their mental models mainly based on reading the text, with the illustration playing a supplementary role. Last but not least, participants' attention was likely to be drawn to irrelevant decorative icons on the illustration which might disrupt the process of constructing a mental model about the concept to be learned.

Hung (2014) found similar attentional distribution when sixth graders read an expository text with illustrations. The participants were instructed to read aloud the material and then retell the content. Two types of dance honeybees use to communicate were described and explained in the material. The text areas included headings, figure captions, and main text, whereas the illustration areas consisted of decorational, representational, and interpretational illustrations. The proportions of these areas were 4, 3, 12, 24, 2, and 12%, respectively (43% of the material was blank space). Overall, 63% of the fixations were made in the text areas (totally 19% in space) and only 27% of the fixations were made in the illustration areas (totally 38% in space). The mean fixation duration in the text areas was about 50 ms longer than that in the illustration areas, whereas the mean saccade length was shorter in the text than in the illustration areas. In the text area, the main text and the captions received more attention (fixation counts and duration) than did the headings. Saccade length was shorter in the former two text areas than in the latter. Although readers exploited the interpretation function of the figure captions, they did not utilize the prediction function of the headings. In the illustration areas, readers paid

the least attention to the decorative illustration. However, they did not pay enough attention to the informational illustration areas so that, in their retelling, they simply repeated *what* was mentioned in the text (e.g., the sun is the reference for the direction of the flowers) without fully understanding *why* (which was depicted in the interpretational illustrations). Furthermore, there was a trend that the more the readers fixated on the illustrations, the higher their comprehension levels were.

Liu and Hou (2011) explored the influence of prior knowledge on comprehending a diagram about plate tectonics. Undergraduate and master's degree students majoring in science first depicted what they knew about plate tectonics on paper. Their drawing and retrospective thinking aloud data were the basis for categorizing their levels of prior knowledge (low, medium, and high). They then read instructional material (a diagram with three AOIs: plates, mantle convection area, and labels) with their eye movement recorded. After reading the material, they had to revise their drawing as the posttest. The eye movement patterns revealed that the high-level prior knowledge group paid more attention to the labels than the low-level group (reflected by longer total fixation durations and higher fixation counts). On the contrary, the high-level group had shorter total fixation times in the mantle convection area than the low-level group. There was no difference in average fixation durations among areas or among groups. The high-level group also had more transitions between labels and back-and-forth scanpaths between labels and the plate area than the low-level group. The three groups were further divided into stable and advanced subgroups according to whether they progressed from pretest to posttest. For both the high- and low-level groups, those who improved (the advanced group) in the posttest had more transitions within the plate area compared to those who did not (the stable group). To summarize, participants with high-level prior knowledge had a text-oriented comprehension process. In addition, those who progressed at the posttest had different transition patterns from those who did not. Presumably, the improvement resulted from the attempt to integrate information.

Ho et al. (2014) also investigated how the prior knowledge levels of undergraduate students influenced their fixation patterns when reading a web page with a paragraph and two diagrams. In the text, the relationship between carbon dioxide concentration and global warming was described in four continuous sections. The first section introduced the greenhouse effect, the second section explained the data shown in the diagrams, the third section described the observation from the diagrams that CO₂ concentration and atmospheric temperature might be highly correlated, and the last section argued that greenhouse gas produced by human industry might cause global warming in the current age. To the left or right of the text area, two diagrams were presented to illustrate the changes in CO₂ concentration (top) and atmospheric temperature (bottom) in the past 200,000 years. Participants' prior knowledge levels were assessed by a paper-and-pencil pretest. Eye movements were then recorded when they read the web page. Overall, all participants spent more time fixating on the text area than on the diagram area. The text area was revisited more often than the diagram area. They also fixated longer on the CO₂ diagram than on the temperature diagram. When comparing two groups of participants, those with higher prior knowledge (PK) generally spent more time

reading the material than those with lower PK. The high PK group had longer total fixation duration on the diagram area, especially in the CO₂ diagram, than the low PK group. The high PK group also revisited both diagrams more often than did the low PK group. In the text area, the high PK group paid more attention than the low PK group did to the middle two sections describing the data in the diagrams and their correlated relationships. Heat maps (representing total fixation duration) showed that the low PK group focused on definition terms in the first introduction section and seldom inspected the diagrams, whereas the high PK group fixated on keywords in the text area and paid attention to the axes in the diagrams. Concerning the transition between the text and diagram areas, the high PK group had more transitions than the low PK group, especially after reading the second and fourth sections. The high PK group also switched between the two diagrams more often than did the low PK group. The results suggest that the two groups used different strategies when reading inquiry-based science material, that is, the high PK group used a text-graphic integration and data comparison strategy while the low PK group used a text-based strategy. It is possible that participants with higher PK had the ability to comprehend and examine the diagrams whereas those with lower PK did not, so that the latter group relied more on reading the text.

Chen and Wu (2012) investigated the effects of coloring on comprehending geometry proofs. Undergraduates unfamiliar with geometry were recruited as participants. The stimuli included a question, a figure, and a worked-out example proof. Equivalent lines or angles on the figure were colored in two of four items. After reading all items, participants were assessed by a paper-and-pencil recall test. The eye movement patterns during reading revealed that total fixation time in the text area was significantly longer than that in the figure area for most items. It should be noted that the text area was larger than the figure area (the proportion of the figure area was 17~30%). When the size of each AOI was taken into consideration, fixation duration per pixel was larger in the figure than in the text area. This may indicate that when reading geometry proofs, the figure area provides critical information for comprehension. The finding that the percentage of total fixation time in the figure area was about 38~48% which was higher than the proportion of area 17~30% also supported this interpretation. Moreover, the effect of coloring was observed on the initial comprehension time in the critical proposition area (CPA) in the worked-out proof, that is, there was shorter initial comprehension time in the colored than in the control conditions. The initial comprehension time was the sum of the durations of the first-pass fixations in the CPA, and subsequent fixations back-and-forth between the CPA and the figure area before leaving this CPA to move to another part of the text area. Coloring the figure area made the critical segments more salient and easier to process after reading the critical proposition in the text. It also facilitated integration between the colored segments and the corresponding CPA, thus resulting in shorter initial comprehension time. However, there was no significant difference in the look-back time and total fixation time between the colored and control conditions (except for shorter total fixation time in the most difficult item in the colored condition than that in the control condition). To summarize, when reading geometry proofs, participants spent more time fixating on

the figure area (taking the number of pixels into consideration) to extract information for comprehension. In addition, coloring the critical segments in the figure area facilitated processing of the critical proposition in the text area and the integration between the CPA and the figure area.

In two studies, Jian, Wu and colleague (Jian and Wu 2012; Jian et al. 2014) investigated the effect of adding arrows for comprehending illustrated material. In their first study, the material was composed of text and an illustration about neural pathways for processing emotional stimuli. Undergraduate and graduate students who were not familiar with neuroscience were recruited as participants. Half of them read the original material without arrows and the other half read the material with arrows indicating the sequence of information processing. After reading the material at their own pace, they were assessed by a comprehension test with text-based, sequential, and integrative questions. Overall, the readers spent about 80 and 20% of total fixation durations on the text and illustration areas, respectively. There was no effect of adding the arrows on the total fixation durations in the different AOIs. However, the arrow group had shorter mean saccade length in the illustration area than the no-arrow group. There were three kinds of scanning patterns, namely, transiting between text and illustration areas at the beginning (60 and 73% in the arrow and no-arrow groups), finishing reading the whole text before inspecting the illustration (5 and 18%), and viewing the illustration before reading the text (35 and 9%). Although a majority of participants in both groups read the text first and then inspected the illustration after reading two or three sentences, the arrow group tended to fixate on the illustration before reading, while the no-arrow group tended to finish reading the text before inspecting the illustration. The effect of adding arrows was also revealed in better performance in the sequential and integrative comprehension questions by the arrow than the no-arrow group. Thus, adding arrows for the sequence of information processing influenced the way participants read the material and their comprehension levels.

In the study of Jian et al. (2014), the learning material was composed of two diagrams about how a flushing cistern works. One of the diagrams depicted the outlet process of flushing while the other illustrated the inlet process, and both of them were presented on the same page on the monitor. The components in the flushing cistern were explained in a separate page before the participants saw the main diagrams; in other words, there was no label on the diagrams to reduce the number of stimuli that may attract participants' attention. Half of the participants received material with additional numbered arrows on the diagrams to illustrate the sequence of the flushing processes. The participants were undergraduate students majoring in non-science subjects so that they were not familiar with the flushing processes. They were instructed to study the diagrams within 5 min and then describe the steps involved in the flushing processes. The results showed that the arrow group had significantly better understanding of the operation steps of the flushing cistern than the non-arrow group. Overall, the arrow group spent slightly more time (181.8 s) than the non-arrow group (158.7 s) on studying the diagrams, but the difference was not significant. However, the mean saccade length was shorter in the arrow than in the non-arrow group. When the two diagrams were analyzed separately, both of them

received longer first-pass fixation times from the arrow than the non-arrow group, but there was no difference in the second-pass fixation time. There was also less saccade switching between the two diagrams in the arrow than in the non-arrow group. Furthermore, Jian et al. conducted sequential analyses of fixation transition among ten AOIs (five components in each diagram). The first-pass transition patterns showed that the arrow group tended to fixate back-and-forth between components according to the numbered arrows, while the pattern was not as apparent in the non-arrow group. The total-pass transition patterns were more diverse than the first-pass ones. However, the arrow group seemed to follow the numbered sequence whereas the non-arrow group tended to compare the same components in the two diagrams. This study illustrated the process of learning kinematic information from static diagrams. Without the aid of the numbered arrows, the participants tended to inspect the same components in both diagrams to infer the operation mechanism. On the other hand, the arrow group benefited from the numbered arrows by following the sequence so that they had better learning outcomes.

To summarize, a general text-directed comprehension process was found. Participants had more and longer fixations in the text than in the illustration areas. However, more attention would be devoted to the illustration if it provided additional information (such as the figure in a geometry proof). Furthermore, participants with more prior knowledge, compared to those with less prior knowledge, were more likely to inspect the illustration and tended to switch between critical parts in the text and illustration areas. Without sufficient prior knowledge or additional arrows, participants may not be capable of extracting information from the illustration. Thus, the only source they could rely on was the text. With the assistance of arrows signaling the sequence of events, participants could better follow the process depicted in the illustrations. Another possibility is that an illustration can be interpreted in multiple ways (such as illustrations in novels and paintings); in this situation, text or captions may constrain the meaning that the authors intended to convey, resulting in text-based comprehension strategies. This speculation may be consistent with the finding of Underwood et al. (2004). In their studies, participants were instructed to verify a statement about a photo. When the photo was presented before the statement, the inspection time on both was longer than when the photo was presented after the statement. That is, when the statement was presented first, it constrained the interpretation of the following photo so that the participants inspected the photo simply for confirmation. However, when the photo was presented without supplementary statements, participants might have to remember several interpretations and some details of the photo so that they spent more time on the photo and verifying these alternatives with the following statement. Undoubtedly, this speculation demands future investigations for confirmation.

13.2.2 Comparing Presentation Modes in Multimedia Material

She and Chen (2009) compared the effects of interaction modes (animation versus simulation) and sensory modality modes (narration versus on-screen text) on

learning the mitosis and meiosis processes by seventh graders. Scores of the posttest and retention-test 5 weeks later were consistent with eye movement measures (total inspection time and mean fixation duration in the pictorial areas). For the animation mode, learning outcome was better with narration than on-screen text, confirming the modality effect. The animation-narration group had longer total inspection time and larger mean fixation duration in the pictorial areas than the animation-text group. On the contrary, for the simulation mode, learning outcome was better with on-screen text than narration, presumably because it was easier to understand the difficult concepts through simulation with permanent text than fleeting narration. Analogously, participants in the simulation-text group paid more attention to the pictorial areas than those in the simulation-narration group. For the on-screen text mode, animation presentation may result in a split attention effect because learners could not control the learning pace. It follows that the animation-text group had shorter inspection time in the pictorial areas and had worse learning outcomes than the simulation-text group. However, although the animation-narration group focused more on the pictorial areas than the simulation-narration group, there was no significant difference in their learning outcomes. To summarize, without control of presentation rate, animation had beneficial effects when accompanied by narration. On the contrary, with the control of learning pace, simulation had facilitatory effects when learners could obtain information from steady text than from uncontrollable narration. In addition, the results also demonstrated that the longer the inspection time, the more effort was devoted to the learning material, and the better the learning outcome was.

Liu et al. (2011) investigated the redundancy effect of presenting both on-screen text and voice-over narration when nonscience-major undergraduate students read web pages with static graphical illustrations of weather systems. Each page consisted of a static graphical illustration and one of three types of verbal information, namely, on-screen text only (picture-text, PT), on-screen text and voice-over narration simultaneously (picture-text-voice, PTV), and voice-over narration only (picture-voice, PV). Eye movements were recorded when they read three web pages. After reading each of the pages, perceived cognitive load was rated separately. The perceived cognitive load was highest in the PTV condition and lowest in the PV condition, whereas that in the PT condition was similar to that in the PTV condition. The total fixation duration was longest in the PT condition, and there was no difference between the PTV and PV conditions (number of fixations showed a similar pattern). The text area was fixated on longer in the PT than in the PTV condition; although there was no difference in the illustration area between these two conditions. Comparing the PT and PTV conditions, the results suggest that the redundancy effect of voice-over narration was marginal; on the contrary, adding voice-over narration even reduced the time spent in the text area. In addition, replacing on-screen text (PT) with voice-over narration (PV) seemed to relieve cognitive load resulting from competition in the visual modality between pictorial and text information (before and when transforming the visual form of text information into verbal form). No difference among conditions in the illustration area was observed in this study.

In the study of She and Chen (2009), inspection time in the pictorial areas increased when the combination of interaction and sensory modality modes were

suitable for learning (i.e., animation-narration and simulation-text). However, in the study of Liu et al. (2011), no difference was observed in the pictorial area among conditions. Instead, adding voice-over narration reduced inspection time in the text area (comparing PTV and PT conditions). It should be noted that the two studies differed in the visualization used. While that in the study of She and Chen it changed dynamically (controlled by the participants or the system), that in the study of Liu et al. remained static. Further and systematic investigation, in which static illustrations, animation and simulation modes together with text or narration are manipulated, may elucidate how learners combine information from different modalities on the one hand, and how verbal information facilitates information extraction from the illustrations on the other.

13.2.3 Real Classrooms

Yang et al. (2013) demonstrated the application of eye-tracking techniques in a real classroom. In their study, an instructor gave a lecture with a PowerPoint presentation concerning evidence of the existence of dinosaurs and theories about the causes of their extinction. Due to the limitations of the current eye-tracking techniques, it was presented as one-to-one instruction. There were various formats in the PowerPoint slides, such as outline, photo only, photo with text, and conceptual graphic and text. Twenty-one undergraduate students participated in this study, with half of them majoring in earth science (ES) and familiar with the issues included in the lecture, whereas the other half majored in other science disciplines (non-earth-science, NES). The major findings are as follows. First, the proportion of time spent on the PowerPoint slide when it was presented (percentage of viewing time, PVT) was about 29.7–59.2%, indicating that the students spent some of the time elsewhere, presumably on the instructor. The PVT on slides with text and photos (43.8%) was higher than that on slides with photos only (35.1%). On the other hand, the average fixation duration on slides with text and photos (261 ms) was shorter than that on slides with photos only (280 ms). Second, concerning different AOIs (title, text, and picture) in a slide, students rarely looked at titles. When both text and pictures were presented in the slide, the proportion of time spent in zone (PTSZ) was higher for text than for pictures, but the average fixation duration was generally higher for pictures than for text. However, for slides with conceptual graphics, the difference between text and pictures decreased. Also, the PTSZ was higher for conceptual graphics than for photos. Third, comparing students with different levels of prior knowledge, ES students tended to spend more time than NES students on text. As for pictures, either an opposite pattern or no significant difference was observed. Detailed analysis of slides with text and conceptual graphics describing the possible causes of dinosaur extinction revealed that, compared with NES students, ES students tended to focus on keyword areas in the text and critical areas in the conceptual graphics. ES students also had more between-zone scan paths. There are multiple sources of information in a real classroom, for instance, text and illustrations on

slides, instruction, and gestures of the instructor, and interaction between students. Learners with different levels of prior knowledge may rely on different sources to gain knowledge. In addition, compared to multimedia learning material with recorded narration (which may be redundant if it is the same as the on-screen text), the instruction from the lecturer may play a different role and serve to guide students' attention to important parts of the slides.

13.2.4 Problem Solving

Tsai et al. (2012) investigated the scanning patterns when undergraduate students majoring in computer engineering solved one multiple choice problem about which of the four factors (i.e., rainfall, slope, debris, and temperature) could cause landslides. Their think-aloud protocols during problem solving were analyzed and served as the basis to categorize them into successful and unsuccessful problem solvers. Overall, fixation durations on the chosen option were longer than those on the rejected options; similarly, the percentage of fixation durations (i.e., fixation durations in proportion to the total fixation durations in the trial) were longer on the relevant than on the irrelevant factors. Sequential analysis of transitions between different AOIs representing factors and the question showed that successful problem solvers had a tendency to transit from irrelevant factors to relevant factors, while unsuccessful problem solvers showed the reverse pattern and tended to look back to the question. The findings imply that some students had difficulty solving the problem, especially distinguishing between relevant and irrelevant factors.

Liu and Shen (2011) examined the strategies used by third and fifth graders when they solved propositional problems concerning concentration. In the orange juice test, for instance, option A consisted of one cup of juice and four cups of water whereas option B consisted of four cups of juice and one cup of water; participants had to decide whether option A or B was sweeter. The orange juice test could be presented iconically (i.e., with colored cups representing juice and water) or symbolically (i.e., with numbers, e.g., 1/4 for option A and 4/1 for option B). The test items also varied in difficulty. Overall, the third graders had longer total fixation duration and longer total time (including fixations, saccades, and blinks) than the fifth graders. Both groups of students had longer average fixation duration when solving problems with symbolic than with iconic representations. In addition, there were three possible transition patterns, namely, transition between cups of juice and water in the same option, transition between the same drink in both options, and cross transition between cups of juice in option A (or B) and cups of water in option B (or A). The third graders most often compared the same drink in both options, and sometimes compared cups of juice and water in the same option. Their transition patterns did not change with variation in presentation types (symbolic or iconic) or item difficulty. However, the fifth graders had different transition patterns that varied with item difficulty and presentation types. Thus, the fifth graders seemed to engage in different strategies flexibly according to the type of problem to be solved.

Lin and Lin (2014b) also investigated the eye movement patterns when senior high school students solved five geometry problems concerning similar triangles, in which an unknown length of a specific side should be calculated. These five problems varied in difficulty, in that they demanded horizontal/vertical translation, reflection, and mental rotation for one of the triangles. The pass rate confirmed the manipulation of item difficulty (at least, the fifth problem was the most difficult item) and it could be predicted by the participants' perceived cognitive load. Eye movement measures (total fixation time, fixation count, and number of runs) followed the order of difficulty. That is, difficult items had longer total fixation time, higher fixation counts, and more runs than easy items. Total fixation time and fixation count also positively correlated with cognitive load. Comparing heat maps of unsuccessful solvers to successful solvers revealed that unsuccessful solvers tended to inspect the whole problem, and had longer fixations in the informational areas (such as digits for lengths of sides of the triangles and the intersection between triangles). As the total fixation time lengthened when the difficulty increased, it follows that unsuccessful solvers had longer inspection times. In addition, they might not be able to distinguish critical regions from less important areas, so they tended to fixate on all locations in the problem.

In another study, Lin and Lin (2014a) examined how senior high school students solved the same five problems with a handwriting device. Thus, the AOIs included a one-line question, a diagram of similar triangles, and the output area (where the solution written through the handwriting device was presented). For the most difficult problem #5, successful solvers spent more time in the output area than the unsuccessful ones. Similarly, the correlations between the perceived difficulty of item 5 and number of fixations as well as total fixation time in the output area were negative, whereas they were positive in the diagram area. That is, when it was difficult to solve a problem, solvers needed to process the diagram extensively. In addition, unsuccessful solvers tended to have more fixations and look backs in the question areas. To summarize the two studies, fixation counts and durations signified the ease and efficiency in problem solving; therefore, they positively correlated with item difficulty. Furthermore, unsuccessful solvers might not be able to distinguish relevant from irrelevant information, so that they fixated on the whole diagram extensively to solve the problem.

Chen et al. (2014) investigated the eye movement patterns when science-major undergraduate students solved physics problems presented in either text or picture formats. The computer-based assessment performance was better with text than with picture presentation. For both formats, correct items received longer mean fixation duration (MFD) and more rereading time in proportion (RRTp) than incorrect items. The mean saccade distance (MSD) was shorter for correct items than for incorrect ones with picture presentation but there was no difference between the correct and incorrect items with text presentation. Comparing text and picture presentations, for both correct and incorrect items, MFD was longer, MSD was shorter, and RRTp was less with picture format than with text format. This finding suggests and confirms that the picture had higher information density than the text (longer MFD and shorter MSD); however, readers had to integrate information from the text rather than from the picture, resulting in more rereading of the text. In addition, the total

inspection time for the text presentations was longer than for the picture presentations. Furthermore, one of the purposes of their study was to investigate whether eye movement measures could predict accuracy in problem solving. The results of the Generalized Estimating Equation (GEE) showed that MFD had the largest positive effect on accuracy (followed by RRT_p) in both formats. When participants spent more time fixating and rereading the problems, they were more likely to answer them correctly. MSD had a negative effect on accuracy with picture presentation but no effect with text presentation. When participants made shorter saccades in solving the problems, they were more likely to have correct answers. In addition, for correct items, participants tended to fixate on critical areas in pictures and keywords in texts. They also fixated on these critical areas longer and reinspected them more often and longer. Thus, longer fixations and more reinspections, especially in the critical areas, increased the likelihood of correctly answering the problems.

To summarize, fixation durations and counts were positively correlated with item difficulty. Younger problem solvers might have few strategies available, whereas older problem solvers could change their strategies flexibly. Similarly, unsuccessful solvers might not be capable of identifying relevant parts of the problem, so that they might fixate on the problem extensively or leave the relevant part for the irrelevant part. On the contrary, Chen et al. (2014) found that mean fixation duration and reinspection frequency positively correlated with performance accuracy. This might seem to be contradictory at first glance; however, fixation duration and frequency are indicators of attention allocation. As mentioned in the first section, an AOI is fixated on and attended probably because it is interesting, relevant, or difficult to process. If enough test items are available in a single study, researchers can further analyze the interaction between item difficulty and accuracy. Maybe more attention is devoted to difficult problems; at the same time, when the problem solver pays attention to the relevant part of the problem, s/he is likely to succeed in solving the problem. Indeed, it was observed in the studies surveyed above that successful solvers tended to fixate on relevant areas longer than unsuccessful solvers. It should be noted that the participants in the study of Chen et al. were undergraduate students majoring in science. Thus, both the ability to identify relevant areas of the problem and whether enough attention is allocated to them contribute to problem-solving performance.

13.3 Summary and Suggestions for Future Investigation

Eye-tracking techniques take advantage of the fact that people naturally move their eyes to sample visual information for further processing. Where the participants fixate, how long they retain their gaze, and the sequence of fixations among several AOIs reveal the cognitive operation underlying information extraction from and processing of the material. The spatial, temporal, and sequential aspects of eye movements are influenced by both the stimulus properties and participants' knowledge about the cognitive relevance of the stimuli. The same stimulus can be processed in different ways if the task demands are different.

The common findings from the 15 articles about science education conducted in Taiwan are summarized as follows (refer to the summary table in the Appendix). First, a general text-oriented comprehension strategy in processing material with text and illustrations was found. Also, there were more fixations and longer inspection time for the text than for the illustration areas. Nevertheless, the general pattern altered when the illustration provided additional information (such as figures in a geometry proof) and when the participants had sufficient knowledge to obtain critical information from the illustration. In multimedia environments, the combination of presentation mode (static illustration, animation, and simulation) and modality of verbal information (text and narration) will influence how learners process the visual representation. Second, participants had different transition patterns among AOIs. Participants with higher-level prior knowledge tended to make transitions among (critical) AOIs, whereas those with lower-level PK rarely inspected different AOIs alternatively. Moreover, older or successful problem solvers had more and flexible transition strategies whereas younger or unsuccessful solvers had less appropriate and fixed strategies. In addition, additional arrows or coloring could guide participants' attention and facilitate the comprehension process. Specifically, participants could follow the sequence of events signified by arrows, and their initial comprehension could be facilitated by coloring relevant parts of the illustration. Third, participants with higher-level PK tended to fixate on the relevant parts of the material longer and more often than those with lower-level PK. However, without sufficient PK, participants may fixate on all of the material thus lengthening their inspection time to solve the puzzle.

Enlightened by the findings of the surveyed studies, a plausible sequence of research can be formulated as follows. Applying eye-tracking techniques to study the learning process and information processing during problem solving can first illuminate how high achievers or experts extract and utilize information, and then examine the difference between low and high achievers. They may differ in the abilities to determine the relevance of the material and to comprehend the material, especially the illustrations. Retrospective interviews may reveal possible sources of the deficiency. Afterwards, an intervention program can be designed accordingly. When the learners have sufficient knowledge of why some parts of the material are relevant and how to process that information, their eye movements can be recorded to assess their improvement. Whether they have similar eye movement patterns as those of high achievers can be an indicator of the necessity of further training.

Nevertheless, it should be noted that eye movement patterns can diverge substantially with different task demands. In addition, if multiple parts of the material are relevant to the task, their saliency may attract different amounts of attention. This may be artifact if saliency is unrelated to the research purpose. Furthermore, total fixation duration and mean fixation duration can have different meanings. For example, participants may process more information during each fixation in one AOI (long mean fixation duration), but they may not devote much time or attention to that AOI (short total fixation duration). Thus, a set of eye movement measures should be considered simultaneously for appropriate interpretation of the data.

Within the past few years, an increasing number of researchers have devoted themselves to applying eye-tracking techniques to educational environments. After basic eye movement patterns (such as attentional distribution between text and illustration areas as well as the strategies used by participants with different abilities) are well documented, fine-grained cognitive processing can be scrutinized. In addition, systematic manipulation of prior knowledge, presentation mode, etc. can elaborate our understanding of how people with different abilities succeed or fail in learning from different formats of material about a variety of content areas.

Appendix

Appendix: Summary of the 15 articles conducted in Taiwan (in order of appearance in the main text)

Note: PK (prior knowledge): High/ Medium/ Low; T: text; P: picture; E: experimental group; C: control group; R: right; W: wrong

Author	Liu and Chuang	Hung	Liu and Hou	Ho et al.	Chen and Wu
Year	2011	2014	2011	2014	2012
Issue	Learning from material with text and illustrations				
Material	Text+picture	Text+picture	Picture+ labels	Text+ pictures	Text+picture
Topic	Earth science (wind, atmospheric pressure)	Biology (bee communication)	Earth science (plate tectonics)	Earth science (greenhouse effect and global climate change)	Math (geometry proofs)
Participants	Undergraduate (nonscience major)	Grade-6	Undergraduate and above (science major)	Undergraduate and above (mostly nonscience major)	Undergraduate and above (not familiar with the subject)
Number of participants	8	6	36	13	31
Participant variables			PK (H/M/L: 10/17/9)	PK (H/L: 6/7)	
Material variables					Coloring (within subject)

Author	Liu and Chuang	Hung	Liu and Hou	Ho et al.	Chen and Wu
Year	2011	2014	2011	2014	2012
Total fixation duration	T>P		Label: H>L Illustration: H<L	T>P T: H>L (sec 2) P: H>L	T>P; T<P (per pixel) E<C (initial comprehension time)
Total fixation count	T>P	T>P	Label: H>L	(Reinspection) T>P T: H>L (sec 3) P: H>L	
Mean fixation duration		T>P	No difference		
Mean saccade length		T<P			
Transition pattern	Refer to the visuals when there was text describing the kinematics of the system		Transitions between labels as well as between labels and plates: H>L	High PK: text and diagram, between diagrams; low PK: rare	
Author	Jian and Wu	Jian et al.	She and Chen	Liu et al.	Yang et al.
Year	2012	2014	2009	2011	2013
Issue	Learning from material with text and illustrations		Comparing presentation modes in multimedia material		Real classroom
Material	Text+picture	Picture	(animation versus simulation)+(text versus narration)	Picture+(text or voice or both)	PPT+instructor
Topic	Neuroscience	Mechanics (flushing cistern)	Biology (mitosis and meiosis)	Earth science (types and causes of thunderstorms)	Earth science (extinction of dinosaurs)
Participants	Undergraduate and above (not familiar with the subject)	Undergraduate (nonscience major)	Grade-7	Undergraduate (nonscience major)	Undergraduate (science major)
Number of participants	42	40	24	16	21

Author	Jian and Wu	Jian et al.	She and Chen	Liu et al.	Yang et al.
Year	2012	2014	2009	2011	2013
Participant variables					PK (H/L: 10/11)
Material variables	Arrows (between subject)	Numbered arrows (between subject)	(animation versus simulation)+ (text versus narration) (between subject)	Text or narration or both (within subject)	
Total fixation duration	T>P	E ~ C first-pass fixation duration: E>C	P: A-N>A-T P: S-T>S-N P: S-T>A-T P: A-N>S-N	All: PT>PTV, PT>PV, PTV ~ PV T: PT>PTV P: PT ~ PTV	Slides with PT>P in one slide: T>P T: H>L P: no difference
Total fixation count				All: PT>=PTV, PT>PV, PTV ~ PV	
Mean fixation duration			P: A-N>A-T P: S-T>S-N P: S-T>A-T	All: no difference T: PT>PTV	Slides with PT<P in one slide: T<P
Mean saccade length	P: E<C	E<C			
Transition pattern	Majority: transiting between text and illustration at the beginning; E: more likely to inspect the illustration before reading; C: more likely to finish reading the text before inspecting the illustration	E: followed the sequence; C: compare same components in both diagrams			H: more transition among AOs
Author	Tsai et al.	Liu and Shen	Lin and Lin	Lin and Lin	Chen et al.
Year	2012	2011	2014b	2014a	2014
Issue	Problem solving				
Material	Picture + question	Iconic versus symbolic	Picture + question	Picture + question	Text versus picture

Author	Tsai et al.	Liu and Shen	Lin and Lin	Lin and Lin	Chen et al.
Year	2012	2011	2014b	2014a	2014
Topic	Earth science (factors causing landslide)	Math (propositional problems)	Math (geometry: similar triangles)	Math (geometry: similar triangles)	Physics
Participants	Undergraduate (computer engineering)	Grades-3 and 5	Senior high school students	Senior high school students	Undergraduate (science major)
Number of participants	6	37	63	63	63
Participant variables	Performance (successful/unsuccessful: 3/3)	Grade (G3/5: 19/18)	Performance (successful/unsuccessful)	Performance (successful/unsuccessful)	
Material variables		Iconic and symbolic presentation (within subject)			Text versus picture (within subject)
Total fixation duration	Chosen > rejected; relevant > irrelevant	G3 > G5	Longer for difficult items and unsuccessful solvers	For difficult problem: picture > output	T > P
Total fixation count			Higher for difficult items	For difficult problem: picture > output; question area: more fixation and look-back (unsuccessful solvers)	(Reinspection) T > P R > W positive correlation with accuracy
Mean fixation duration		Symbolic > iconic			T < P R > W positive correlation with accuracy
Mean saccade length					T > P P: R < W P: negative correlation with accuracy
Transition pattern	Successful: irrelevant → relevant; unsuccessful: relevant → irrelevant and back to the question	G3: single transition pattern; G5: changed their transition pattern among problems			

References

- Chen, C.-Y., & Wu, C.-J. (2012). Eye movements during geometry proof reading: Text contrasting with figure and the colored effects. *Journal of Educational Practice and Research*, 25, 35–66. [Chinese]
- Chen, S.-C., She, H.-C., Chuang, M.-H., Wu, J.-Y., Tsai, J.-L., & Jung, T.-P. (2014). Eye movements predict students' computer-based assessment performance of physics concepts in different presentation modalities. *Computers & Education*, 74, 61–72.
- Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice* (2nd ed.). London: Springer.
- Ericsson, K., & Simon, H. (1993). *Protocol analysis: Verbal reports as data* (2nd ed.). Boston: MIT Press.
- Findlay, J. M., & Gilchrist, I. D. (2003). *Active vision: The psychology of looking and seeing*. New York: Oxford University Press.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717–742.
- Henderson, J. M. (2003). Human gaze control during real-world scene perception. *Trends in cognitive sciences*, 7, 498–504.
- Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16, 850–856.
- Ho, H.-N. J., Tsai, M.-J., Wang, C.-Y., & Tsai, C.-C. (2014). Prior knowledge and online inquiry-based science reading: Evidence from eye tracking. *International Journal of Science and Mathematics Education*, 12, 525–554.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford: Oxford University Press.
- Hung, Y.-N. (2014). “What are you looking at?” An eye movement exploration in science text reading. *International Journal of Science and Mathematics Education*, 12, 241–260.
- Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Amsterdam: Elsevier.
- Jian, Y.-C., & Wu, C.-J. (2012). The effect of arrows in an illustration when reading scientific text: Evidence from eye movements and reading tests. *Chinese Journal of Psychology*, 54, 385–402. [Chinese]
- Jian, Y.-C., Wu, C.-J., & Su, J.-H. (2014). Learners' eye movements during construction of mechanical kinematic representations from static diagrams. *Learning and Instruction*, 32, 51–62.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329–354.
- Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1196–1206.
- Kaakinen, J. K., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *The Quarterly Journal of Experimental Psychology*, 64, 1372–1387.
- Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., et al. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115.
- Lin, J. J.-H., & Lin, S. S.-J. (2014a). Tracking eye movement when solving geometry problems with handwriting devices. *Journal of Eye Movement Research*, 7(1):2, 1–15.
- Lin, J. J.-H., & Lin, S. S.-J. (2014b). Cognitive load for configuration comprehension in computer-supported geometry problem solving: An eye movement perspective. *International Journal of Science and Mathematics Education*, 12, 605–627.

- Liu, H.-C., & Chuang, H.-H. (2011). An examination of cognitive processing of multimedia information based on viewers' eye movements. *Interactive Learning Environments, 19*, 503–517.
- Liu, C.-J., & Hou, I.-L. (2011). Explore the influence of prior knowledge on understanding in scientific diagrams through eye tracking. *Bulletin of Educational Psychology, 43*, 227–250. [Chinese]
- Liu, C.-J., & Huang, C.-F. (in press). Innovative science educational neuroscience: Strategies for engaging brain waves in science education research. In M.-H. Chiu (Ed.), *Science education research and practice in Taiwan - Challenges and opportunities*. Springer.
- Liu, C.-J., & Shen, M.-H. (2011). The influence of different representations on solving concentration problems at elementary school. *Journal of Science Education and Technology, 20*, 621–629.
- Liu, H.-C., Lai, M.-L., & Chuang, H.-H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior, 27*, 2410–2417.
- Martinez-Conde, S., & Macknik, S. L. (2007). Windows on the mind. *Scientific American, 297*, 56–63.
- Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*, 205–231.
- Radach, R., Huestegge, L., & Reilly, R. (2008). The role of global top-down factors in local eye-movement control in reading. *Psychological research, 72*, 675–688.
- Radach, R., & Kennedy, A. (2004). Theoretical perspectives on eye movements in reading: Past controversies, current issues, and an agenda for future research. *European Journal of Cognitive Psychology, 16*, 3–26.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology, 62*, 1457–1506.
- Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*, 219–226.
- Rayner, K., Pollatsek, A., & Reichle, E. D. (2003). Eye movements in reading: Models and data. *Brain and Behavioral Sciences, 26*, 507–526.
- Rayner, K., Miller, B., & Rotello, C. M. (2008). Eye movements when looking at print advertisements: The goal of the viewer matters. *Applied Cognitive Psychology, 22*, 697–707.
- Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.
- Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience, 7*, 361, doi:10.3389/fnhum.2013.00361.
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*, 1–21.
- She, H.-C., & Chen, Y.-Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers & Education, 53*, 1297–1307.
- Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17.
- Tatler, B. W. (2009). Current understanding of eye guidance. *Visual Cognition, 17*, 777–789.
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education, 58*, 375–385.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *Quarterly Journal of Experimental Psychology, 57A*, 165–182.

- Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99.
- Yang, F.-Y., Chang, C.-Y., Chien, W.-R., Chien, Y.-T., & Tseng, Y.-H. (2013). Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computer & Education, 62*, 208–220.