Monitoring Children's Learning through Wearable Eye-Tracking: The Case of a Making-Based Coding Activity

Michail N. Giannakos, Sofia Papavlasopoulou, Kshitij Sharma
Norwegian University of Science and Technology - NTNU

***Abstract*—Learning activities for/with children include rich interactions with peers, tutors and learning materials (in digital or physical form). During such activities, children gain new knowledge and master their skills. Automatized and continuous monitoring of childrens learning is a complex task, but, if efficient, can greatly enrich teaching and learning. Wearable devices, such as eye-tracking glasses, have the capacity to continuously and unobtrusively monitor childrens interactions, and such interactions might be capable of predicting childrens learning. In this work we set out to quantify the extent to which childrens gaze, captured with eye-tracking glasses, can predict their learning. To do so, we collected data from a case study with 44 children (8-17 years old) during a making-based coding activity. Our analysis shows that childrens gaze can predict their learning with 15.79% error. Our results also identify the most important gaze measures with respect to childrens learning, and pave the way for new research in this area.**

Introduction

Learning gains are currently assessed using pre- and post-tests. However, such tests pose additional strain on children, do not provide to-the-minute information, require considerable effort and language competence, and sometimes are not appropriate (e.g., for children with special abilities). Wearable eye-tracking devices have made it possible to monitor subtle phenomena, such as the quality of social interactions, mental health and work-outs [16]. However, despite the great potential of wearable eye-tracking devices to enable the continuous and unobtrusive monitoring of learning, research in this direction remains rather underexplored.

In this work, we argue that if eye-tracking data can accurately infer the learning gain from standardized pre- and post-tests, then learning can be measured continuously and unobtrusively from children, as well as providing a reliable solution when conducting a test is not possible.

Thus, to investigate the potential of wearable eye-tracking devices to monitor learning, we formulate the following research questions:

1) Can wearable eye-tracking devices be utilized to predict childrens learning during a making-based coding activity?
2) What are the most important gaze-based predictors of childrens learning?

To tackle the aforementioned research questions, we conduct a study in which we use eye-tracking glasses to capture childrens gaze during a making-based coding activity. The data is coupled with a standardized pre- and post-test. We then apply machine learning to investigate the

 © 2019 IEEE

possibility of inferring childrens learning from the gaze data. By investigating the feasibility of childrens gaze informing us about their learning progress, we provide a path towards a technology design that enables the continuous and unobtrusive monitoring of their learning and overcomes the disadvantages of traditional standardized tests.

Related work

Previous works have employed several practices to assess childrens learning. A common practice is to collect the actual artefact (code, in the case of programming) created in childrens projects and then analyse it using a framework, rubric or checklist (e.g., [17]). A more qualitative way of assessing childrens learning is to conduct interaction analysis [7]. Interaction analysis techniques investigate human activities, such as talk, non-verbal interaction, and the use of artefacts and technologies, identifying routine practices and problems and the resources for their solution. Another qualitative way of assessing childrens learning is to use the think-aloud approach [9]. Think-aloud can reveal a more authentic cognitive process, but it increases the cognitive load and often distracts the participant from the core task [3]. The most common, and more subjective and quantitative [9], method to capture childrens learning gain is to utilize standardized post-tests, multiple-choice instruments, or quizzes that quantify childrens learning [6].

Capturing childrens eye movements (gaze) when they look at a stimulus while working on a task can provide a range of useful information. Gaze data is a good proxy for various cognitive mechanisms [13], and data on childrens gaze provides insight into their visual attention, which enables researchers to explore childrens cognitive processes. Existing studies have applied eye-tracking methods to investigate program comprehension and debugging [9] [3]; however, these studies have concentrated on the source code rather than looking at other representations, with the different representations being taken to be different areas of interest (AOIs). In addition, related work [10] has focused on university students or even professionals, and has utilized stable eye-trackers, thus being able to eye-track only on-screen activities. Although this practice might provide accurate results for professionals or adult students (since the coding/learning activities occur only on-screen), in the case of children, focusing on on-screen activity alone captures only part of the learning experience and provides inaccurate insights. In our study, we utilized eye-tracking glasses in order to collect eye-tracking data to capture the gaze of children while they were working with instructional tasks (1) on the screen, (2) while talking with other children and (3) while working with other objects/materials outside the screen.

In short, standardized pre- and post-tests provide reliable assessments of childrens learning, but at the same time entail certain disadvantages. For instance, children need to invest time and effort into these tests, and it is not possible to apply the tests to every activity. In addition, assessments that involve heavy text can be inappropriate for young students (e.g., those in primary education) who have not yet fully mastered the written word or have certain disabilities and other learning difficulties. Thus, if we can automatically and pervasively collect gaze data that can accurately predict childrens progress (as assessed from standardized tests), then childrens learning can be measured continuously and unobtrusively using wearable eye-tracking devices.

Experiment: Case Study of a Making-Based Coding Activity

Participants

The study was conducted in a dedicated lab space at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. Over a two-week period, 44 children from 3rd to 12th grade (aged 8–17 years old) participated in a coding activity. The sample comprised 12 girls (mean age: 12.64, SD: 2.838) and 32 boys (mean age: 12.35, SD: 2.773). Five workshops took place over two weeks, that followed a specially designed coding activity (described below). Our activities were organized for children who were novices at coding and who participated voluntarily. Once the participants had been selected, a researcher contacted their teachers and parents in order to obtain the necessary consent from the legal guardian for data collection. Consent from the children themselves was also obtained.

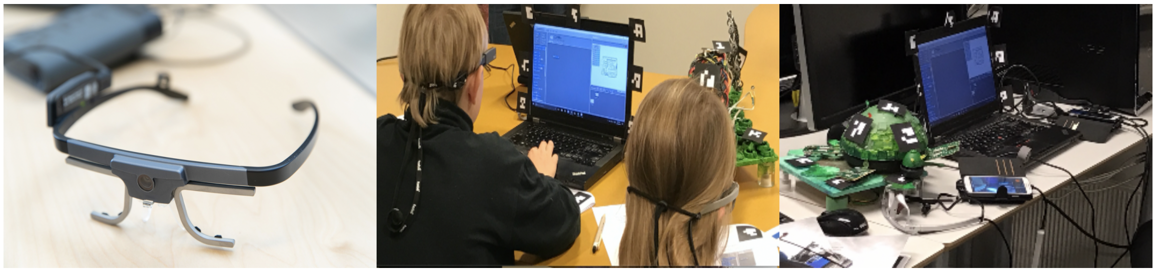

Figure 1. Left: Eye-tracking glasses; Middle: Snapshot from the study; Right: QR codes in the main AOIs.

Apparatus

In order to collect childrens gaze data during the coding activity, we utilized eye-tracking glasses that allowed us to track childrens on-screen and off-screen gaze (see **see Figure 1, left, middle**). We used SMI RED 250 and TOBII mobile eye-trackers working at 60Hz. A sampling rate of 60Hz is considered sufficient for usability studies. In order to automatically compute the necessary features from our data, we put QR codes in the main AOIs (i.e., screen and robot), with the exception of the face AOI, which was automatically extracted via face recognition (**Figure 1, right**).

Activity

The workshop activities were based on the constructionist approach, as one of the main principles of this is learning-by-making. The workshop was conducted in a largely informal setting, as an out-of-school activity, and lasted for four hours in total; it was designed for children without (or with minimal) previous experience in coding. Participants in the workshops comprised various student groups, ranging from 8-17 years old, who were invited to rooms at NTNU that have been specially designed for creative purposes. The children worked collaboratively in triads, and student assistants supported them during the activity, with each assistant observing and helping one or two teams. The childrens gaze was monitored during the activity.

The activity consisted of two parts (**Figure 2, left**): first, developing interaction concepts in digital robots using the Arduino hardware platform, and second, developing games using Scratch (a block-based programming language). Specifically, for the first part Arduino was attached to the digital robots to connect them to the computer. At that point, an extension of Scratch called Scratch for Arduino (S4A) provided the extra blocks needed to control the robots. Children had to accomplish a series of simple loops to make the robots react to the physical environment (i.e., each robot had a specific movement that caused lights to be turned on and off; see examples in **Figure 2, right**). The duration of this part varied for each team, and lasted between 45 minutes and 1.5 hours, ending with a break before the next section.

The Scratch programming language uses colourful blocks grouped into categories (motion, looks, sound, pen, control, sensing, operators, and variables), with which children can develop stories, games, and various types of animation (see **Figure 3**). During the second part of our workshop, the children developed their own games by collaboratively designing and coding using Scratch. They created their games step by step by iteratively testing and coding them, using loops and other computational thinking concepts. Depending on the needs of each team, complex programming concepts were introduced by the assistants according to the relevance to the teams project. This section lasted approximately three hours.

Research Design

The children completed pre- and post-knowledge acquisition tests. These consisted of nine questions of increasing difficulty, and were adopted from literature on how to assess childrens computational thinking skills [6] (see an example in **Figure 3**). The children took approximately 10 minutes to finish the tests. The tests were paper-based and were manually graded by the researcher. Children wore the eye-tracking glasses during all parts of the activity, and their gaze was

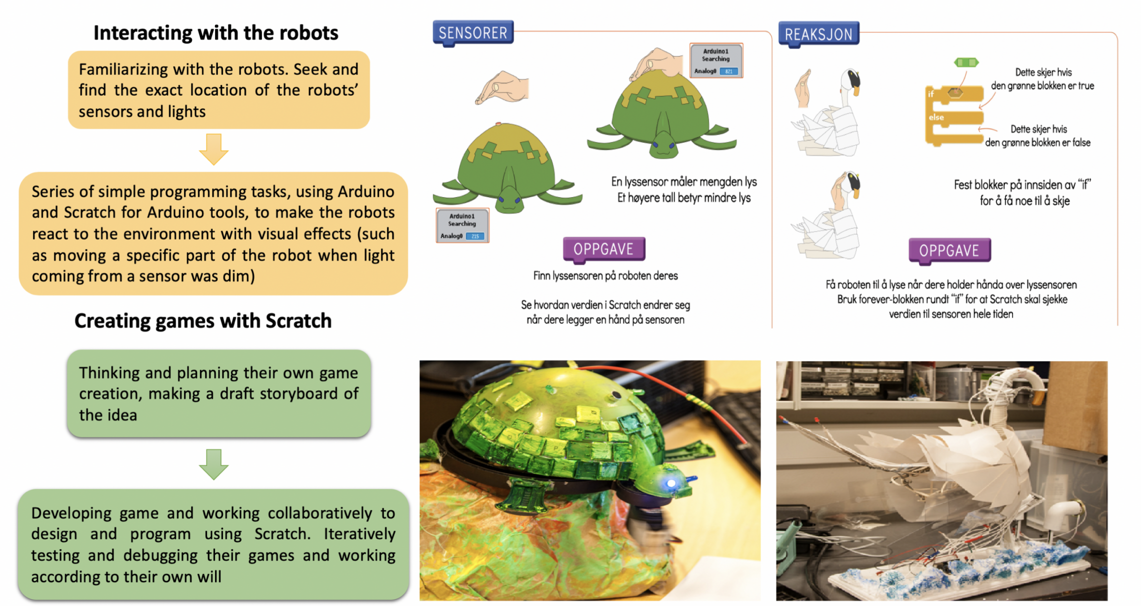

Figure 2. Left: Description of the two parts of the activity; Right: Examples of the tasks (top) and the robots children (bottom).

continuously and unobtrusively captured.

Measures

First, we calculated the relative learning gain (RLG) [5]. RLG is more accurate compared to learning gain, since it takes into consideration the difficulties inherent in gaining more knowledge if the learner is already very knowledgeable in a subject. In this work, RLG is considered as the dependent variable.

$$RLG = \begin{cases} \dfrac{Posttest - Pretest}{Max.\,in\,pretest - Pretest}, if\ Posttest \geq Pretest \\ \dfrac{Posttest - Pretest}{Pretest}, if\ Posttest < Pretest \end{cases}$$

We employed 15 measures (**Table 1**), and the observations of the researchers during the study; these comprised four AOIs, six transitions (to and from one AOI to another) and five behavioural measurements. All the measurements were calculated for a fixed time window of 10 seconds and then aggregated for the whole session. For variables such as transitions, we counted the frequencies and normalized them by the duration of the window; variables including fixation duration and skewness of saccadic velocity were averaged over the time window and later aggregated.

For the cognitive load, the mean and SD of pupil diameter was combined with the number of long fixations and saccade length in order to provide a reliable measurement of cognitive load (to counterbalance the effect of emotions). In particular, to measure the cognitive load of participants we calculated the four measures proposed by Buettner [2] (i.e., mean pupil diameter (PDM), pupil diameter SD (PDS), number of fixations longer than 500ms (NLF) and saccade speed (SS), see the following formula). These values were calculated and combined for every participant.

$$Cognitive\ load = \begin{cases} \{\mathbf{0}\ if\ PDM \leq median\ (PDM) \\ \mathbf{1}\ if\ PDM > median\ (PDM)\} \\ + \\ \{\mathbf{0}\ if\ PDS \leq median\ (PDS) \\ \mathbf{1}\ if\ PDS > median\ (PDS)\} \\ + \\ \{\mathbf{0}\ if\ NLF \leq median\ (NLF) \\ \mathbf{1}\ if\ NLF > median\ (NLF)\} \\ + \\ \{\mathbf{1}\ if\ SS \leq median\ (SS) \\ \mathbf{0}\ if\ SS > median\ (SS)\} \end{cases}$$

To account for variations in pupil diameter due to susceptibility of pupil diameter measures towards time of the day, age and gender, after removing noise we used the first two minutes of the data from each participant to normalize their pupil diameter. Further, to account for changes in

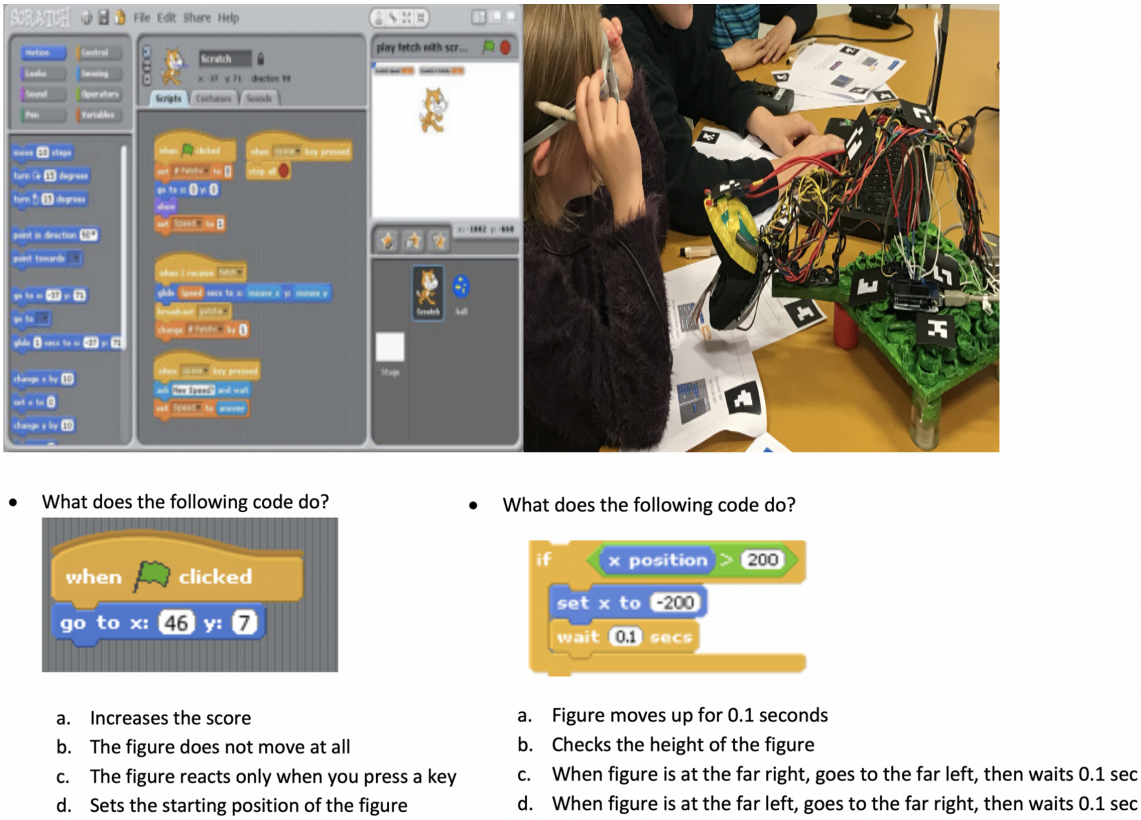

Figure 3. Top-left: Interface of Scratch IDE; Top-right: Children interacting with the robots; Bottom: Example of questions used in the pre and post-tests.

brightness, pupil diameter was normalized within the brightest and the darkest frames in the field of view for each child. For attention, we utilized the average fixation duration. Fixation duration is context sensitive and has been associated with mind-wandering; however, in the context of programming [1] it has been attributed to higher levels of information processing and attention (in terms of time spent on the content of a specific part). For anticipation, the skewness of the saccade velocity histogram how fast the eyes were moving was used. Anticipation is related to the individuals familiarity with the interface (stimulus). If a person is familiar with a given interface (Scratch IDE, in our case), the eye movement between the different parts of the interface to solve a particular problem should be faster compared to that of a person who is not familiar with the interface. The saccade velocity skewness gives us a general idea this phenomenon, since the higher the skewness of the saccade velocity histogram the higher the overall speed of eye movements [14].

Analysis

To test the potential difference in gaze behaviour between children with low versus high RLG, we formed two different groups via a median split on RLG. Subsequently, an independent samples analysis of variance (ANOVA) was conducted between children with low and high RLG.

To identify how gaze data features can predict RLG, we divided the whole data-set into training and testing subsets, retaining data from 10 children for testing. We perform a 10-fold cross-validation (retaining 10 children for testing each time) to remove sampling bias from the training set. We used ensemble learning with model trees to predict the childrens RLG using the gaze data features. Ensemble learning uses multiple learning algorithms to obtain better predictive performance than could be obtained from using

Measure	Name	Type	Meaning
Proportion of time not looking at -screen, -face and -robot.	Nothing	AOI	Off-task activity.
Proportion of time looking at the robot.	Robot	AOI	Interaction time with the robot.
Proportion of time looking at faces.	Face	AOI	Interaction time with other students and the tutor.
Proportion of time looking at the screen.	Screen	AOI	Interaction time with the IDE.
Transition from face to screen.	Face.screen	Transition	Shift of attention between a face and the screen.
Transition from screen to face.	Screen.face	Transition	Shift of attention between the screen and a face.
Transition from other children's/tutor's faces to something else.	Face.	Transition	Shift of attention between a face and another AOI.
Transition from screen to something else.	Screen.	Transition	Shift of attention between the screen and another AOI.
Transition from something else to the screen.	.screen	Transition	Shift of attention between another AOI and the screen.
Transition from something else to other children's/tutor's faces.	.face	Transition	Shift of attention between another AOI and a face.
Mean and S.D. of pupil diameter; number of long fixations; saccade length.	Cognitive Load	Behavioral	Child's mental effort.
Average fixation duration.	Attention	Behavioral	Child's level of focus.
Skewness of saccade velocity histogram (how fast the child's eyes are moving).	Anticipation	Behavioral	Child's level of anticipation.
Blink rate.	Fatigue	Behavioral	Child's tiredness.
Gaze similarity (common gaze): The proportion of time all children were looking at the same AOI within a time window of 4 seconds.	Joint Attention	Behavioral	Level of collaboration between the team of children.

Table 1. Summary of the gaze measures used in our study.

any of the constituent learning algorithms alone. We used the normalized root mean squared error (NRMSE) metric; NRMSE is the proposed metric for prediction models regarding learning [12].

Results

On average, childrens learning increased from the pre- to the post-test, with the majority of children having positive RLG (except nine that had a gain of 0 i.e., scored the same in the pre- and the post-assessment) after the coding activity (mean RLG: 0.45, SD: 0.38). A bi-modality test on RLG (Hartigans dip test: $D = 0.11$, $p < .001$) revealed that the data were not unimodal and thus we could perform a median split between low and high RLG. To investigate any potential differences between the two groups, we performed an ANOVA including childrens RLG as a dependent variable and their gaze behaviour as independent variables. All statistical analyses reported had a significance level of 0.05. As we can see from the outcome data in **Table 2**), the time the children looked at the screen and at the face had an impact on their RLG, as did the behavioural measures of cognitive load, joint attention, anticipation and attention.

Utilizing wearable eye-tracking devices to capture childrens learning

To identify how the different eye-tracking features can predict RLG, we divided the whole data-set into training and testing sets, with data from 10 children retained for testing. We performed a 10-fold cross-validation (again retaining 10 children for testing each time) to remove sampling bias from the training set. The average value across all cross-validation folds for the testing sets of NRMSE was 15.79%. To identify whether we could make a reliable prediction with a smaller portion of our data, we attempted to predict RLG using different data segments (e.g., 100%, 50%, 25%). With 75% of the data-set, NRMSE was 18.45%; with 50% it was 19.68%; and with 25% it was 21.26% (using 25% of data translates to dividing the data into four quarters based on length of time and using only the first quarter for prediction purposes). Thus, when we moved to 25% of the data-set, we noticed a significant increase in NRMSE.

Important gaze-based predictors of children's learning

Regarding the role of each of the 15 features (see Variable importance for the RLG column in Table 2, the three most important (i.e., important predictors) were (1) childrens anticipation (i.e., how fast the childs eyes moved) during the activity; (2) interaction time with the IDE (i.e., proportion of time the child was looking at the screen); and (3) the childs level of focus (i.e., childs attention average fixation duration). The three least important features were (1) child looking at something other than the main AOIs (i.e., not at the screen, face or robot); (2) interaction time with the robot (i.e., proportion of time the child was looking at the robot); and (3) childs transition from something else to other childrens/tutors faces.

Insights from observations during childrens learning

From the qualitative observations made during the study, we also identified that the time spent in the various regions of Scratch interface differed between the low and the high performers, but also between the younger and the older children. Children with better final games (based on the submitted code) shared certain characteristics: They started coding from sprites and changing the costumes of their characters, and all members of the team were engaged with coding and spent a lot of time looking at the screen to familiarize themselves first with the commands. In addition, they had a lot of interaction with the assistants. Another interesting insight coming from the observations is that children older than 13 spent more time on the commands, output and scripts compared to the younger ones, who spent more time on the sprites. In general, younger children focused more on the aesthetics of the characters they were designing for their games.

Discussion

Using gaze data to understand and predict learning is not a new approach (see a recent literature review by Lane and DMello [8]). Previous studies have revealed significant findings in terms of gaze data and their appropriateness for modelling learners engagement with an activity and predicting learning. However, previous results have focused on digital environments [8], thereby completely missing the authentic learning that occurs outside the screen, and have focused only on adult learners in lab studies. The present study, in contrast, considers how continuous and unobtrusive monitoring of childrens gaze during an authentic and rich activity can help us to predict their learning. Our results suggest that gaze data produced during the learning activity can be good predictors of childrens learning, with the screen AOI and some behavioural features serving as very good predictors.

Variable (Based on names in Table 1)	Mean (SD)		ANOVA			Normality		Variable importance* for the RLG
	Low RLG	High RLG	F	p	Cohen's d (effect size)	W	p	
AOIs								
Nothing	0.04 (0.02)	0.06 (0.02)	0.26	.60	0.1	0.94	.41	0
Robot	0.16 (0.05)	0.16 (0.06)	1.2	.28	0.3	0.97	.43	5.25
Face	0.12 (0.03)	0.14 (0.03)	5.21	.02	0.7	0.93	.11	18.52
Screen	0.39 (0.08)	0.33 (0.11)	4.54	.03	0.6	0.94	.08	92.61
Transitions								
Face.screen	0.12 (0.02)	0.13 (0.02)	0.1	.85	0.1	0.94	.45	29.09
Screen.face	0.13 (0.01)	0.12 (0.02)	1.6	.20	0.3	0.93	.13	17.25
Face.	0.13 (0.02)	0.13 (0.01)	0.3	.57	0.1	0.95	.56	48.45
Screen.	0.14 (0.01)	0.13 (0.02)	1.7	.20	0.3	0.91	.41	31.89
.screen	0.12 (0.02)	0.14 (0.02)	0.45	.54	0.1	0.91	.45	11.22
.face	0.12 (0.01)	0.13 (0.01)	0.39	.53	0.1	0.96	.22	11.18
Behavioral								
Cognitive Load (scalar between 0 and 4)	1.51 (1.11)	2.14 (1.05)	3.75	.05	0.5	0.90	.18	41.37
Attention (milliseconds)	310.10 (162.52)	228.64 (85.51)	4.72	.03	0.6	0.98	.11	76.94
Anticipation (degrees/ms)	0.41(0.46)	0.02 (0.60)	5.95	.01	0.7	0.93	.12	100
Fatigue (mm, normalized between 0 & 1)	0.62 (0.08)	0.64 (0.08)	0.54	.45	0.2	0.92	.09	22.82
Joint Attention (scalar between 0 & 1)	0.46 (0.03)	0.39 (0.15)	4.89	.03	0.6	0.96	.22	52.33
* The importance is scaled between 0 (least important, should not be confused with not important) and 100 (most important).								

Table 2. Differences in gaze behaviour between children with low versus high RLG, and variable importance for RLG.

Design considerations

In the context of programming, attentional distribution and shifts (i.e., AOI and transitions) have been found to relate to performance [11] [10]. In our study, transitions between the different AOIs and the non-core AOIs (i.e., nothing and robot) were found to be insignificant, while measurements related to the core AOI (i.e., screen) were found to be significant (positively). This result is not surprising, and confirms the relationship between on-task engagement and learning performance. Thus, it is important to cultivate interaction quality time with the IDE when designing coding activities for children. However, this is also difficult, since it is extremely easy for children to disengage (as we also observed in our study).

The most important results come from the measures related to behaviour in particular, anticipation, attention and joint attention were found to be the most important predictors of childrens learning. Anticipation is a key factor for expertise; experts are known to have a top-down approach to solving problems, with the main factor of such processing being that experts gaze is driven by their experience/knowledge [15]. Prior knowledge provides children with a working hypothesis, and thus they are able to anticipate. In this study, some children had prior knowledge about the programming environment (some also developed knowledge quickly during the activity); thus, including the measure of anticipation allowed us to encode the expertise of the children. The fact that the measure of anticipation was found to be the most important predictor, and was also significantly higher in children with high RLG, supports related work in the context of programming, but also informs future work about the importance of utilizing anticipation-related features when monitoring childrens learning and skills development.

Attention and joint attention were also found to play an important role in childrens learning. In particular, moments of joint attention are a viable predictor of the quality of collaboration and outcome [5]. Thus, when designing learning activities and technologies for children it is important to enhance childrens collaboration with one another, keep their attention/focus on the task at hand, and support anticipation during interactions. These considerations might sound simple, but they are directly connected to several approaches employed in childcomputer interaction (CCI), such as using interactive elements to increase sociability or game elements to increase attention, or setting the physical space design in a way that amplifies childrens abilities to collaborate and stay focused.

Interestingly, childrens cognitive load was not a top predictor of learning. Cognitive load is a widely used measurement in the intersection of eye-tracking and learning. This result indicates the importance of relaxation of mental effort during learning, and is grounded in the importance of meaningful learning. For meaningful learning to occur, the cognitive process/load of the activity should be moderate; or, as Csikszentmihalyi [4] demonstrated, achieving the optimal or flow experience that neither frustrates nor deters the learner. This can be achieved by selecting relevant information, organizing it into coherent mental representations, and integrating it with prior knowledge, but also by helping children to relax by designing the activity in a way that mixes mentally heavy with playful and social elements.

Working with children in wearable eye-tracking device studies

The present study is one of the first to use mobile eye-tracking data to examine childrens learning gain. When the children were initially asked to wear the glasses, they were curious to know more about them; therefore, we had to spend time explaining why we were running the experiment and how the gaze data would look on the monitor. Most of the children had never seen anything similar before and were excited to wear the glasses, without complaining about them being heavy or uncomfortable. However, they were less tolerant than adults with respect to the temperature of the glasses, which heat up after being in use for some time; after informing us of this, we had to take a break and allow the children to remove the glasses for a while. In addition, there were practical difficulties inherent in capturing childrens gaze because of their constant head movements during the activity, and the fact that the eye-tracking glasses were designed mainly for adults (and thus sized for adult heads, causing irritation). Moreover, the collaborative concept of the activity encouraged children to heavily interact with each other, which made it difficult to collect high-quality data. However, our study proved that it is feasible to eye-track children using glasses, since 75% of the data were of good quality and gave us very interesting insights regarding gazes ability to predict childrens learning in rich activities. Overall, when conducting studies in which children are asked to wear eye-tracking glasses it is important to look carefully at the value and need for childrens involvement in the development process, and of the value and purpose of childrens participation.

Despite the aforementioned difficulties, we found that contemporary wearable eye-tracking solutions (SMI and Tobii glasses) have the capacity to be used with children as end-users. This does require some additional work from the researcher, as there is a need to pay extra attention to monitoring the activity and ensuring proper calibration, when needed (e.g., removing glasses during play), and proper use of the glasses. However, besides those considerations, eye-tracking glasses are a good solution that enables us to unveil rich interactions in the context of CCI.

Another important consideration when working with children in wearable eye-tracking device studies pertains to appropriate preparation of the apparatuses and the space. Eye-tracking glasses, in contrast to stable eye-tracking, need a carefully prepared and modelled environment (e.g., AOIs) that allows the researcher to automatize the post-processing, but at the same time endures the extensive playing, learning and socializing activities that occur during a collaborative learning activity with children.

Limitations

The findings support our proposition that childrens gaze has the capacity to infer their gains during a learning activity, but are subject to certain limitations. The participants of our study were children; this represented an appropriate sample for our study, since we wanted a child population that could effectively participate in a typical coding activity, use the eye-tracking glasses, read the standardized test and provide accurate responses. However, younger or older populations (e.g., kindergarten, university students, etc.) might produce slightly different results. The generalizability of our findings is somewhat constrained by the design of the learning activity, since longer, more passive or less collaborative activities might give slightly different results. However, this study employs a learning design that is widely used (i.e., collaboratively coding using a visual programming tool) and can thus serve as a baseline for future studies.

Pre- and post-tests were used to provide the ground truth; these tests consisted of standardized questions based on the literature [6]. Different assessment techniques, such as interaction analysis [7] or think-aloud [9], could have allowed us to assess childrens cognitive process in a more authentic manner; however, the subjective quantification of learning would had been challenging. Despite its limitations, standardized pre- and post-tests represent a reliable method for measuring childrens learning [6]. Complex constructs, such as cognitive load and attention, could have been measured by asking the participants themselves (e.g., via a NASA–TLX survey or a dual task), either at the end of the task or at regular intervals during the task. However, such methods are neither automatic nor pervasive, do not provide temporal measures, add inherent complexity of the main task and cause important disruptions to the task and collaboration. Thus, to compute the behavioural constructs (Table 1), we employed gaze measures from the literature [2] [1] that have been found to provide adequate results in similar contexts (e.g., programming).

Conclusions

Overall, our work shows that monitoring childrens gaze with eye-tracking glasses holds the potential to intuitively capture their learning progress. We provide evidence that leveraging eye-tracking glasses can be a viable method to accurately track childrens learning during rich learning activities. In summary, the contribution of this paper is twofold: (1) we conducted an in-the-wild study that provided data on childrens gaze during a making-based coding activity, and quantified the capacity of childrens gaze data to give accurate predictions of their learning; and (2) we identified the importance of the various gaze-based data to predict childrens learning, and discussed how these findings can inform future CCI and learning technology research.

Future work should collect data from different learning activities. Cross-validating and extending our findings will allow us to build generalized prediction models and identify the learning activities in which we can most accurately predict childrens engagement. This will allow us to build an integrated understanding of the potential (as well as the limitations) of using childrens gaze data to infer their learning progress.

Acknowledgment

The authors would like to express their gratitude to all of the children, teachers and parents for volunteering their time. Our very special thanks go to Letizia Jaccheri, Kristin Susanne Karlsen, Ioannis Leftheriotis, Amanda Jrgine Haug, Lidia Luque Fernandez, Marjeris Sofia Romero, Eline Stenwig and Kristoffer Vens Monsen.

The project has been recommended by the Data Protection Official for Research, Norwegian Social Science Data Services (NSD), following all the regulations and recommendations for research with children. This work was funded under the European Commissions Horizon 2020 COMnPLAY-Science project (Project Number: 787476) and the Norwegian Research Council under the projects FUTURE LEARNING (number: 255129/H20) and Xdesign (290994/F20).

REFERENCES

1. R. Bednarik, and M. Tukiainen, "Visual attention tracking during program debugging," *Proc. third Nordic conference on Human-computer interaction*, pp. 331–334, 2004.
2. R. Buettner, "Cognitive workload of humans using artificial intelligence systems: Towards objective measurement applying eye-tracking technology," *Proc. KI 2013: Advances in Artificial Intelligence*, pp. 37–48, 2013.
3. T. Busjahn, C. Schulte, B. Sharif. et al., "Eye tracking in computing education," *Proc. Tenth Annual Conference on Computing Education Research*, pp. 3–10, 2014.
4. M. Csikszentmihalyi, *Flow and the Psychology of Discovery and Invention*, New York: Harper Collins, val. 56, p. 107, 1996.
5. P. Dillenbourg, S. Lemaignan, M. Sangin, N. Nova, and G. Molinari, "The symmetry of partner modelling," *I. J. Computer-Supported Collaborative Learning.*, vol. 11, no. 2, pp. 227–253, 2016.
6. S. Grover, S. Cooper, and R. Pea, "Assessing computational learning in K-12," *Proc. 2014 Conference on Innovation Technology in Computer Science Education*, pp. 57–62, 2014.
7. B. Jordan, and A. Henderson, "Interaction analysis: Foundations and practice," *Journal of the Learning Sciences.*, vol. 4, no. 1, pp. 39–103, 1995.
8. H, C, Lane, and S. K. DMello, "Uses of physiological monitoring in intelligent learning environments: A review of research, evidence, and technologies," *In Mind, Brain and Technology*, pp. 67–86, 2019.
9. Y. T. Lin, C. C. Wu, T. Y. Hou et al., "Tracking students cognitive processes during program debuggingAn eye-movement approach," *IEEE Transactions on Education.*, vol. 59, no. 3, pp. 175–186, 2016.
10. U. Obaidellah, M. Al Haek, and P. C. H. Cheng, "A survey on the usage of eye-tracking in computer programming," *ACM Computing Surveys.*, vol. 51, no. 1, pp. 5, 2018.
11. S. Papavlasopoulou, K. Sharma, M. Giannakos, and L. Jaccheri, "Using eye-tracking to unveil differences between kids and teens in coding activities," *Proc. 2017 Conference on Interaction Design and Children*, pp. 171–181, 2017.
12. R. Pelnek, "Metrics for evaluation of student models," *Journal of Educational Data Mining.*, vol. 7, no. 2, pp. 1–19, 2015.
13. D. C. Richardson, and R. Dale, "Looking to understand: The coupling between speakers and listeners eye movements and its relationship to discourse comprehension," *Cognitive Science.*, vol. 29, no. 6, pp. 1045–1060, 2005.
14. A. C. Smit, and J. A. M. Van Gisbergen, "A short-latency transition in saccade dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades," *Experimental Brain Research.*, vol. 76, no. 1, pp. 64–74, 1989.
15. T. Van Gog, F. Paas, and J. J. Van Merrinboer, "Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data," *Applied Cognitive Psychology.*, vol. 19, no. 2, pp. 205–221, 2005.
16. R. Wang, et al., "Tracking depression dynamics in college students using mobile phone and wearable sensing," *in Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 1, pp. 43, 2018.
17. J. Yu, and R. Roque, "A survey of computational kits for young children," *Proc. 2018 Conference on Interaction Design and Children*, pp. 289–299, 2018.

Michail Giannakos is a professor of interaction design and learning technology at the Department of Computer Science, and leads the LearnerComputer Interaction lab, at Norwegian University of Science and Technology (NTNU). His research interests centre on developing new ways for humans to interact with interactive learning systems. Giannakos is a member of the executive board of the IEEE Technical Committee on Learning Technology and a senior member of IEEE. Giannakos is also the recipient of a Marie Curie fellowship and the Norwegian Young Talent award, and is one of 20 outstanding academic fellows of NTNU. Contact him at michailg@ntnu.no.

Sofia Papavlasopoulou is a PhD student in the Department of Computer Science at NTNU. Her research focuses on creative programming experiences for children, and the use of emerging technologies and making activities to enable entertaining, engaging and efficient learning. She received her MSc in informatics from the Department of Informatics at the Ionian University, Corfu, Greece. Contact her at spapav@ntnu.no.

Kshitij Sharma is a post-doctoral researcher at the Department of Computer Science at NTNU. His research focuses on applied machine learning, HCI and learning technologies, with heavy emphasis on groups behaviour and physiological data such as eye-tracking, EEG and facial expressions (theoretical and practical methods in digital interaction). He received his PhD in computer science from EPFL. Contact him at kshitij.sharma@ntnu.no.