Analysis of students' eye movements during solving multiple-choice scientific literacy test

Martina KEKULE, Alžběta KREJČÍ

Charles University, Faculty of Mathematics and Physics, Department of Physics Education, V Holesovickach 2, 180 00, Prague, Czech Republic

Abstract. Our aim is to show the possibility of using eye tracking to assess some aspects of the validity of multiple-choice tests. Specifically, we focus on nine tasks with textual and graphical representations from the TOSLS science literacy test. Twenty-three high and undergraduate school students participated in the study. We can assess validity based on quantitative analysis (e.g., the time students take to solve a task, the time they spend on an alternative or a stem, etc.). We can also assess validity based on qualitative analysis of gaze plots, for example we can identify distracting graphical elements, difficult or illegible words.

The measurement of various skills and competences is an integral part of physics education and is often done through tests, especially multiple-choice tests. In this paper, we attempt to show the possibility of using eye tracking to assess some aspects of the validity of such tests. Specifically, we focused on the TOSLS science literacy test [1]. The test contains 28 questions either in text-only version or with graphs. Some of them contain actual newspaper excerpts, either in text or graphical representation. It is not only this large variation in graphical material that can affect the validity of individual items. By using eye movements, we not only get the final product of the student's thinking, i.e. their answer, but we can also observe the process of their thinking. Both quantitative and qualitative analysis of eye movements can help us determine where students are directing their attention and whether they are distracted by something we don't want to measure.

Method

Each of 23 students solved 9 task from the TOSLS test (in the national (Czech) language [2]). These are multiple-choice tasks. Three tasks are text-only and six include graphs: either within a stem (four tasks) or within an alternative (two tasks). The slide with the task was followed by a slide about how confident students were in their answer. Specifically, they selected an option on a five-point scale (from “I was confident” to “I guessed”). After the eye tracking session, we conducted an in-depth interview with a student about their records. A TX300 eye-tracker (300 Hz) and an IVT filter (Tobii Studio 3.2) were used. The minimum fixation duration was set to 60 ms.

Results

We present some preliminary results in the abstract, see Table 1. Almost all tasks were easy for our students, with the exception of Task 21. Students spent between 50 s and 120 s (average total fixation time) solving the tasks and on average made between 189 and 513 fixations per task. It took them the most time to solve tasks 6 and 11. Task 11 contains a larger excerpt from the news within the task stem. Task 6 contains a graph in the task stem. As can be noticed, for all problems with a graph in a task stem and for task 11, students spent on average less time on the alternatives. This is in line with our expectation that for these tasks students would pay more attention to the task stem. Further, as part of the validity assessment, we can look at the attention students paid to the alternatives and assess how attractive the distractors are. Finally, we can look at the students' distribution of attention to the area of a graphically complex graph. For example, although students

engaged with the complex graph in Task 21, they were not overly distracted by unimportant graphical features (see Figure 1).

Table 1. Average total fixation duration (in s) and average fixation count spent on each task.

		total fixation duration /s		fixation count			
		Mean	St. Dev.	Mean	St. Dev.	*	**
Text-only representation	task 1	81.87	26.923	347	107.793	70	91
	task 8	53.71	15.952	238	72.774	96	74
	task 11	112.54	42.208	513	182.989	70	36
A graph within task stem	task 6	117.68	37.971	470	162.011	74	21
	task 18	59.86	22.603	243	80.885	78	26
	task 21	55.86	19.844	244	70.4	52	35
	task 28	74.23	18.89	333	85.438	74	40
Alternatives in graphical representation	task 2	51.76	21.431	189	72.526	74	75
	task 15	58.06	30.811	245	120.518	65	74

* % of students, who answered correctly
** ratio (in %) of total fixation duration mean on alternatives to the whole task area

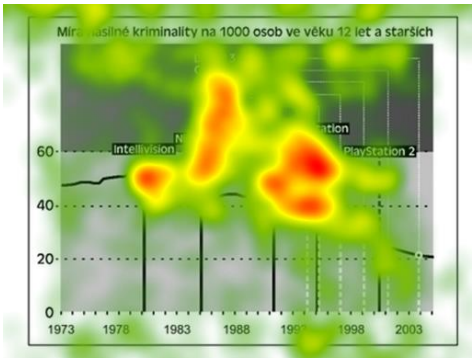

Figure 1. Heat map based on fixation duration of all participants on graph presented in task 21.

Discussion & Conclusion

Our goal was to demonstrate the feasibility of using eye tracking to assess some aspects of the validity of multiple-choice TOSLS test items. Validity can be assessed by quantitative analysis (e.g., the time it takes students to solve the task, the time they spend on an alternative or a stem of the task, the number of fixations on various areas of tasks). We can assess how attractive the distractors are and, for example, the least attractive distractor can also be omitted or replaced. Based on a qualitative detailed analysis of gaze plots, we can identify distracting graphical elements, difficult and illegible words, etc.

References

[1] C. Gormally, P. Brickman, M. Lutz, (2012) Developing a Test of Scientific Literacy Skills (TOSLS): Measuring Undergraduates' Evaluation of Scientific Information and Arguments. *CBE life sciences education* **11** (2012) 364-377. https://doi.org/10.1187/cbe.12-03-0026

[2] *TOSLS test.* Czech version of the test. Available on-line: https://www.physics.ujep.cz/cs/materialy-pro-ucitele. [Accessed:15-March-2024].