DETERMINING THE RELEVANCE OF AUTOMATIZED FEEDBACK ON STUDENTS’ E-TEST RESULTS IN SUPPORTING TEACHERS TO PROMOTE COMPONENTS OF SCIENTIFIC LITERACY IN THE CLASSROOM

T. Rosin, K. Vaino, R. Soobard, M. Rannikmäe

University of Tartu (ESTONIA)

Abstract

Standardised e-testing can be meaningfully integrated within formative assessment strategies through teacher feedback – first through gaining feedback about his/her students’ learning gains and needs and based on that, modifying instruction and secondly, providing supportive feedback to students to facilitate their learning.

The aim of this study is to explore the relevance of the format of an automatic feedback system on portraying students’ e-test results from a science teacher perspective. The automatic feedback is designed so as to describe student gains regarding each scientific literacy (SL) component, while indicating the achieved level. The e-test is initially administered to 9th grade students (15-16 years old, N=1617) and then followed up with semi-structured interviews with 5 voluntary science teachers as a pilot study, related to whether and how they felt that the gained feedback can assist them in modifying their teaching and enhance support for their students’ future learning. The interviews were conducted and analysed using qualitative content analysis. As a result of the study, it was concluded that the format of feedback was generally understandable and meaningful for the teachers.

Keywords: Scientific Literacy, Automatic Feedback, E-testing.

1 INTRODUCTION

The development of students’ competences, to enhance scientific literacy (SL) through school education, has become an important international priority ([1], [2], [3] [4],). Attained competences allow students, for example, to interpret scientific information, create links between knowledge from different disciplines, conceptualize and apply science knowledge in a new situation. Moreover, SL relates to personal attitudes and using critical and creative thinking while searching for a possible solution to a complex problem. Development of these competences has been a priority for educators and are incorporated in science curricula. Although innovative pedagogical approaches and development of aforementioned competences can be central in teaching science according to the Estonian national curriculum ([5], a recent study showed that Estonian students’ problem solving and decision making skills are not strong ([6]). This can be attributed to the fact that the emphasis in science education in Estonia has been on teaching and assessment of lower level cognitive skills ([7]).

Sadler ([8]) points out that feedback is received or provided by two main audiences, teachers and students. From the teacher perspective, feedback has been described as the process in which teacher receives student-related information in order to use it as the basis for altering instruction ([9]). On the other hand, through gathered feedback, the teacher is able to provide supportive feedback to students to facilitate their learning ([10], [11], [12]).

Based on the earlier, a model providing feedback to both student and teachers leading to a science e-test has been developed in Estonia according to which students’ gains in a range of SL components can be determined. In addition, it can be a means to support science teachers in gaining specific feedback about the performance of a particular student group and enabling them to address the gaps in students’ skills and learning needs that need to be paid more attention in the future.

Therefore, the aim of this study was, as a first step, to explore the relevance of the format of the externally developed standardized feedback on students’ e-test results, from a science teacher perspective. For that aim, the following research question was posed:

Does the developed format of automatic feedback on students’ e-test results provide meaningful support for science teachers for promoting SL in the classroom?
2 METHODOLOGY

The methodology consists of two components. In the first part, the model used for the developing of the science e-test and automatic feedback, is described. In the second part, the conduct of the qualitative study to obtain relevant feedback for science teachers, is presented.

2.1 The science e-test model

The science e-test model is based on a set of contextual test items which include subject-specific (chemistry, physics, biology and geography) cognitive items (including both, computer-assessed and open-ended manually assessed responses) in an interdisciplinary context and identifies levels of students’ knowledge and skills, related to the perceived attributes promoting scientific literacy (SL). Such SL attributes are based on cognitive components within the curriculum and are encoded in the model as follows:

Component 1 – Component 4 (from now on: component = C) subject-related scientific knowledge and skills;
C5–C6 socio-scientific decision-making and scientific data handling abilities;
C7–C10 components of science inquiry learning.

The model allows assessment at 3 levels according to the complexity of thinking processes exhibited by 9th grade students in undertaking the test item – see Table 1.

2.1.1 Test design

The three levels were arbitrarily labelled as baseline, average and excellence. Each level of each SL component was assessed through two different test items to raise the reliability of the results while mean scores for each item pair were calculated. The test was piloted and improved twice based on the gained results with an aim to achieve a better fit to the predetermined cognitive complexity criteria as presumed by the model (reported elsewhere).

Table 1. Description of three hierarchical levels for three exemplary codes.

<table>
<thead>
<tr>
<th>Code of component</th>
<th>Cognitive component areas</th>
<th>Level I: baseline</th>
<th>Level II: average</th>
<th>Level III: excellence</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Explains and analyses natural objects, phenomena, processes, and the causal relationship between them.</td>
<td>Explains phenomena, or processes.</td>
<td>Explains phenomena, or processes and the causal relationship between them.</td>
<td>Associates different phenomena and processes with each other; explains causal relationships between them.</td>
</tr>
<tr>
<td>C5</td>
<td>Makes decisions on science related, everyday issues; justifies decisions made</td>
<td>Makes a decision on a given issue, taking only one aspect into consideration, while justifying the decision made.</td>
<td>Makes a decision on a given issue, taking more than one aspect into consideration, while justifying the decision made.</td>
<td>Makes a decision on a given issue, takes multiple relevant aspects (e.g. scientific, ethical, moral, economic, personal) into consideration, while justifying the decision made.</td>
</tr>
<tr>
<td>C8</td>
<td>Formulates a research question/hypothesis.</td>
<td>Selects a relevant question/hypothesis from a given list.</td>
<td>Formulates a relevant research question/hypothesis.</td>
<td>Improves the wording of a given research question/hypothesis. (e.g. Fig. 2)</td>
</tr>
</tbody>
</table>

2.1.2 Test administration

As a pilot study, the test was administered to 9th grade students (15-16 years old, N=1617) in March 2019. Students performed the test on personal computers with performance time permitted up to 120 minutes.
2.1.3 The developed format for the automatic feedback

The developed format of automatically generated feedback was designed in such a way so as to describe student gains regarding each SL component, while also indicating the achieved level. In this study, three SL components (Table 1) were chosen from C1-C10 as illustrative examples. Based on the test results, a graph was generated for each student group, taught by a particular interviewed teacher showing the percentage of students achieving each indicated level (baseline, average, excellent) for a specified SL component. Fig. 1 shows student group results regarding the three SL component areas, determined by the test.

![Graph showing student group results](image)

Figure 1. A student group results regarding the three SL component areas determined by the test.

2.2 Interview

Following administration of the e-test and automatic or centrally organised assessment of students' open-ended responses, semi-structured interviews were held with five purposefully selected science teachers from the schools taking part in the pilot study. Table 2 shows the specialty, gender, and working experience of the participants.

<table>
<thead>
<tr>
<th>Teacher /code</th>
<th>gender</th>
<th>specialty</th>
<th>working experience (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Teacher 1</td>
<td>F</td>
<td>physics / biology</td>
<td>12</td>
</tr>
<tr>
<td>Teacher 2</td>
<td>F</td>
<td>chemistry</td>
<td>15</td>
</tr>
<tr>
<td>Teacher 3</td>
<td>F</td>
<td>chemistry</td>
<td>12</td>
</tr>
<tr>
<td>Teacher 4</td>
<td>F</td>
<td>physics/biology/chemistry</td>
<td>16</td>
</tr>
<tr>
<td>Teacher 5</td>
<td>F</td>
<td>geography</td>
<td>12</td>
</tr>
</tbody>
</table>

2.2.1 Interview procedure

The interviews were conducted with each teacher individually. Interviews were supported by the previously generated feedback in the format of computer slides but also as printouts, about the test results of the students taught by an interviewee. The semi-structured interviews were based on the following questions:

1. Was the format of feedback understandable for you? Could you explain?
2. If yes, then what benefits do you see from the feedback on e-test results when you plan your further teaching? (If no, could you explain?)
The interviews lasted about 40 minutes and were conducted in spring 2019. All interviews were recorded and afterwards, transcribed.

The interviewed teachers were able to familiarise themselves with the description of SL components in which their students (on average) achieved at the excellence level, but also in which competences their students’ skills needed to be improved. The assessment of each level of each SL component was illustrated for a teacher by an exemplary test item. Also, the interviewed teachers were able to compare their students’ results with the average results by students’ results with the average results by students within the country, Fig. 2.

Figure 2. Exemplary test item together with a student group results demonstrated to a teacher during interview

2.2.2 Data analysis

The interviews were analysed using qualitative content analysis ([13]). First, the transcribed text was carefully read and repeatedly reread while marking out all meaning units (teacher statements directly related to the research question) from the text. From the meaning units, a number of condensed meaning units were developed which were categorised according to their content.

Example

Meaning unit: “Actually, the tasks enabling to train these skills would be highly needed and the students could do these tasks in the same e-environment.”

→ Condensed meaning unit: “Lack of relevant materials to train students.”

→ Category: Aspects that make it difficult to promote SL in classroom

3 RESULTS

The results are organised according to the interview questions:

Was the format of feedback understandable for you? Could you explain?

Generally, all teachers were able to understand the format and the meaning of the graphs as given in Figures 1 and 2. Also, the description of SL component and in which components their students performed well or need improvement, was generally comprehensible for the interviewees. Still, Teacher 1 was confused about the way the students’ percentages of each achieved level, were presented - the same students who were able to solve an average level task, but also an excellence level task, or respectively, tasks from all levels of a particular SL component, were represented in
both/all percentage bars simultaneously (the bars representing students’ percentages were not mutually exclusive):

I would like to see exactly how many students do I have at the level I, how many at the level II and how many at the level III. (Teacher 1)

Most teachers (Teacher 1, 2, 3, 5) found that the description supporting the graphs was very necessary in order to understand the meaning of the assessed SL component. Teacher 3 even expressed herself as follows:

Without t/he description, next time, I probably wouldn't even bother to look at the numbers in graphs.

Still, Teachers 1 and 2 would have liked to see a more detailed information about the performance of a particular student to be able to provide him/her an individual scaffolding (Authors’ remark: within the next e-test, students will be provided with the automatically generated individual feedback). Teacher 1, 2, 3, 4 also liked the way how the results of their own students were juxtaposed with the state results as a “black and white evidence about how good they were”. On the other hand, as expressed by the Teacher 4:

It is good to see the state-level results, because, for example, if problem-solving skills are generally low across the country, then it is probably not the “fault” of a concrete student or a teacher.

It helped, at least for Teacher 4, to find an explanation to the rather poor performance of her students in problem-solving tasks.

If yes, then what benefits do you see from the feedback on e-test results when you plan your further teaching? (If no, could you explain your answer?)

It was evidenced by all teachers, that the provided feedback made them to analyse the results of their students. All teachers agreed that in case their students participate in the next science e-test, they definitely would like to get detailed feedback about the achievement level according to a particular SL component, not only the final scores of the test. Teachers 1, 3, 4 brought out that the current format of the feedback enabled them to understand which skills need to be promoted more in the future. Teacher 3 explained:

If I see that science knowledge components are generally very good, but for example, decision-making is not, then I will pay probably more attention to that from now on.

Teachers 2, 3, and 4 pointed out, that the feedback helped them to become more aware about the competences assessed by the test, as its format was pretty novel for them. Therefore, regarding some SL components, the students were not even prepared for (Teacher 1, 3, 5). Although most of the them (1, 3, 4) fully agreed that the SL components measured by the test together with the developed test items were age appropriate, meaningful and relevant while integrating knowledge from different subject areas in an every-day life context, the teachers tended to be sceptical about the resources that would enable them to develop the same competences by their students in the future. Teacher 4 said:

It takes a lot of time to develop a meaningful learning task that would promote such aspects. It would be good to have a repository of such tasks somewhere.

Still, teachers 2 and 4 pointed out that during the last years, it was already possible for them to find relevant materials for promoting various SL components by students. But even they found that more materials would be needed. Teacher 5 even emphasised that amongst all these competences, we should not forget factual knowledge without which it is not possible to make connections. Teacher 3 and 4 also pointed out that the existing educational situation together with state exams have not been fully supportive towards promoting wider SL components (e.g. every-day life related problem-solving and decisions making) in science classroom. While the teaching of science subject knowledge and skills has been highlighted for years if not centuries as the only target, it seemed to become difficult for some teacher (e.g. Teachers 2, 5) to broaden the focus and learning goals.

4 CONCLUSIONS

Based on the interview results, it was found that for the teachers, it was important to get detailed feedback on their students’ e-test results regarding the SL components assessed and that the
feedback format was generally understandable. In this aspect, the feedback format fulfilled its purpose. The teachers tended to appreciate both, the numerical information, but also the more detailed evaluative description about the achieved level of their students according to each SL component. Also, the comparison of results between a particular student group with the whole grade 9 student cohort in the country, was seen as useful, although not always felt comfortable by some of the teachers. The teachers found different explanations why some results were low by their students, starting from a lack of teaching resources and preparation time and reaching to the whole school system not supporting enough the novel approach - at the same time, it seemed, not to fully recognising their own role as educational agents eliciting change in classroom.

The teachers found that the current format of feedback was useful as it helped to operationalise the aspects that are currently highlighted in Estonian science education as very often, teachers do not see how to achieve the general goals posed by curricular documents. At the same time, some of the teachers tended to put too much emphasis to the “training” aspect while talking how to promote SL in classroom. Rather it was interpreted by these teachers as how to improve test results next time than how to develop wide spectrum of competences by the students, which, amongst the other methods of assessment, can be assessed by the e-test. Therefore, it was found by the authors, that more emphasis should be put in the future on communicating the purposes of the current assessment format (which is both, summative and formative) and the current highlights in science education amongst the Estonian science teachers. The study also indicated that there is a need to provide science teachers with in-service courses about the formative feedback and how summative aspects of the e-test can be meaningfully integrated with formative (assessment for learning).

REFERENCES