



# ATS STEM

## Assessment of Transversal Skills in STEM



## National Research Report Finland

Iida-Maria Peltomaa

Lachlan Paterson

# 1. INTRODUCTION

## 1.1. THE NATIONAL CONTEXT

### 1.1.1 Description of the structure of the country's education system and levels of education by age

In Finland, participation in early childhood education and care is a universal right for all children under school age, that is, from birth until six years of age. Moderate, means-tested fees are payable. Compulsory pre-primary education starts one year before basic education at the age of six. Municipalities must provide a minimum of 700 hours of pre-primary education per student per year. Pre-primary education is free of charge.

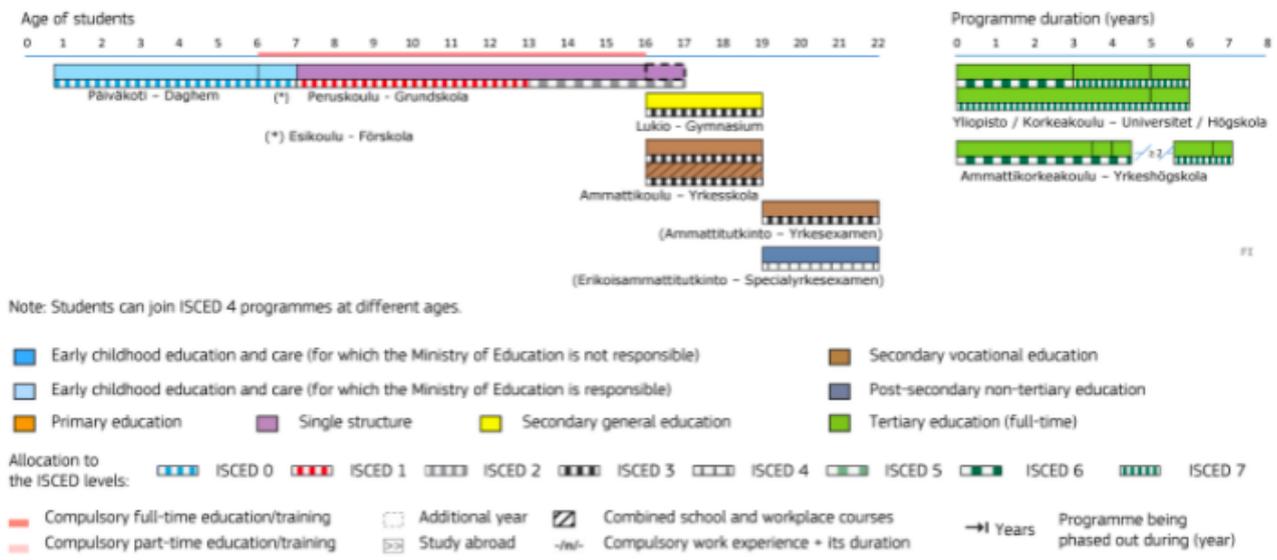
Compulsory schooling begins at the age of 7 and continues for 9 years (grades 1-9). It is provided in a single structure system called basic education. The compulsory subjects in basic education include biology, geography, ethics, physics, chemistry, history, social studies, arts, domestic science, crafts, physical education, mathematics, music, health education, foreign languages, religion, Finnish language and literature, environmental science. In addition to the obligatory subjects, students can study optional subjects. All schools follow the national core curriculum for basic education which allows for local variations, and therefore individual schools and teachers have a lot of freedom in designing their own curricula and instruction.

Upper secondary education is provided by general and vocational upper secondary institutions. The standard age to participate in upper secondary studies is from 16 to 19 years. In 2021, the Finnish government passed a law raising the minimum school-leaving age from 16 to 18 years. Higher education is provided by universities and universities of applied sciences, the former being more academically-oriented and the latter more professionally-oriented institutions. ISCED 8 level qualifications, such as doctorates, can only be granted by universities.

Education is publicly funded, and education from pre-primary to higher education is free. Only two per cent of pupils in compulsory education attend schools that have a private provider, and these schools are also publicly funded. For all students in preschool, basic school and upper secondary education, meals and learning materials are provided free of charge. The differences between schools are small.

Schools are largely free of external controls such as school or textbook inspections. The first national examination takes place at the end of general upper secondary education. The most important quality assurance mechanism is the self-evaluation carried out by the education providers themselves.

Figure 1: Structure of the national education system (Eurydice 2020/21)



### 1.1.2. STEM Education in the national education system

The Finnish curriculum for early childhood education, preschool, basic school, or high school do not mention STEM education. Instead, The Finnish National Core Curriculum for Basic Education introduces the concept of interdisciplinary learning modules that integrate concepts and methods of several subjects<sup>1</sup> into a learning period. “In order to safeguard every pupil’s possibilities of examining entities and engaging in exploratory work that is of interest to the pupils”, all schools must provide at least one interdisciplinary learning module per academic year for every student. Interdisciplinary learning modules provide an opportunity for the teachers to implement STEM education. The curricula for early childhood education, preschool, and high school similarly each enable the implementation of STEM education.

In recent years, the Finnish National Agency for Education has funded some projects related to STEM and STEAM education. Partly due to these projects, and partly due to active teachers, there are some local networks of teachers interested in STEM education. Additionally, LUMA Centre Finland, a science education network of Finnish universities, supports the life-long learning of teachers working in all levels of education from early childhood through to university, and strengthens the development of research-based STEM teaching. The aim of the network is to inspire and motivate children and youth into STEM.

<sup>1</sup> “All subjects are in turn involved in implementing the learning modules as required by the current topic” (National Core Curriculum for Basic Education 2014).

### **1.1.3. Digital assessment in the national education system**

The Finnish National Core Curriculum for Basic Education (2014) requires students “to practise using digital tools for assessing their learning” in the subjects of mathematics, Finnish language, foreign languages, and crafts.

The Finnish National Agency of Education instructs the education providers ie. the local municipalities to ensure that digital assessment tools that contain personal information must be handled with attention to the General Data Protection Regulation (2018) (Chapter 1, Article 4). The digital tool must have a security level that guarantees data protection. The Finnish state authorities do not validate or certify information systems, and therefore the Finnish National Agency of Education cannot take a stand on the information security of any specific information systems, even if they would be widely used in schools.

## **1.2. CONTEXT OF THE PILOT SCHOOLS**

In Finland, a total of nine basic education schools took part in the ATS STEM project. Four of these schools were public primary schools (grades 1-6, ages 7-12), and five of these schools were public lower secondary schools (grades 7-9, ages 13-15). A total of 410 students were involved in the project, consisting of 188 girls and 222 boys. A total of 31 teachers were involved in the project, 20 women and 11 men. 12 different subjects were taught during the project (physics, mathematics, chemistry, biology, geography, Finnish language, arts, crafts, science, physical education, domestic science, English). In total, 22 STEM projects were implemented in the context of the project.

## **1.3. DESCRIPTION OF CASE STUDY SCHOOLS**

### **1.3.1. Individualised description of case study schools**

Both case study schools are public basic schools. In case study school A, there are 13 teachers and 200 students in grades 0–6 (ages 7-12). In case study school B, there are 34 teachers and 430 students in grades 0–9 (ages 7-15).

Both case study schools are located in a sparsely populated rural area, about 30 minutes drive from the city centre of Oulu, North Ostrobothnia, Finland. Oulu is the most populous city in northern Finland, and the fifth most populous in the country. The area was consolidated into the municipality of Oulu in 2009, due to a decrease in the local population. Most of the schools in the area have been closed down due to declining pupil numbers and indoor air problems. However, the village-like local community of the case study schools has grown in recent years, with a relatively large number of eco-spirit families with children having moved to the area. Nature is close by, and there are good opportunities for recreation.

Case study school A operates in a new multi-purpose building, and the school actively cooperates with many outside actors, such as the local association of the 4H youth organisation. Case study school B is active in the Eco-Schools programme, which aims to empower children to drive change and improve their environmental awareness.

Both case study schools have had previous experience in STEM education, and both schools already systematically utilise interdisciplinary and exploratory learning methods, even outside of the ATS STEM project. Both case study schools also have an ongoing collaboration with Fab Lab Oulu, a digital manufacturing working area operating within the University of Oulu.

Teacher A from school A has completed a 25 ECTS STEAM training program organized by the University of Oulu. Teacher B from school B has participated in several short-term in-service trainings about STEM, and as a “tutor teacher” spends 4-6 hours of their weekly working time mentoring other teachers in the municipality of Oulu about STEAM and digitality.

### 1.3.2. Detailed description of the STEM projects developed during the pilot

The case study schools utilised Qridi as the digital tool. It is a digital pedagogical tool developed in Finland that aims to support active learning, and to build student agency and ownership. It allows students to set goals, give out tasks and prompt reflection. Qridi can help students to visualise their learning process by compiling all learning activities, including goals, tasks and reflections into one software. More information is available from their website: <https://qridi.com/edu>. All students of the case study schools had some prior experience using Qridi, but nevertheless all students received training in how to use Qridi at the beginning of the project.

#### 1.3.2.1 Case study school A

Table 1: Project information for school A

|  |   |
|--|---|
| Project started (date of the first lesson) | 15.4.2021   |
| Project ended (date of the last lesson)    | 27.5.2021   |
| Duration of lessons                        | 45 min  |
| Total amount of lessons                    | 28 (of which 10 observed)   |
| Number of students                         | 30  |
| Year level of students                     | 5   |
| Teacher(s) responsible                     | 1 classroom teacher   |
| Subjects involved                          | Environmental Studies, Finnish language, Mathematics, Social Studies, Visual Arts                                     |
| Targeted transversal STEM competencies     | <ul style="list-style-type: none"> <li>● Collaboration</li> <li>● Self-regulation</li> <li>● Communication</li> </ul> |
| Digital tool(s) utilized                   | Qridi   |

|                                       |   |
|---------------------------------------|---|
| Formative assessment strategies       | <ul style="list-style-type: none"> <li>● Sharing learning intentions</li> <li>● Clarifying success criteria</li> <li>● Questioning</li> <li>● Classroom discussions</li> <li>● Giving feedback</li> <li>● Using feedback</li> <li>● Self-assessment</li> <li>● Peer assessment</li> </ul> |
| Artefacts produced during the cycle 1 | Design plans and constructed models of heart and lung function  |
| Artefacts produced during the cycle 2 | Student videos showcasing the activity stations related to the senses that the students constructed during the learning cycle. The final artefacts are intended to be shared with younger students who are invited to play and learn with the activity stations.                          |

Table 2: Learning cycles 1 and 2 for school A

| Learning cycle | Step   | Explanation  |
|----------------|--|--|
| 1              | Defining a real world problem related to SDG's | <p>Exploring different ways to maintain one's own well-being and health. Considering a healthy life for different bodies. Researching and learning basic concepts and searching for information about them.</p> <p>Learning to understand how the human body is made up of different structures and systems, learning in general about bones and muscles, the brain and nervous system, and digestion in particular.</p> |
| 1              | Finding solution(s)                            | <p>What humans are like and how the human body works?</p> <p>The topic was introduced by utilising a method of "non-fiction book clubs". Several lessons were used for getting acquainted with the topic in small groups.</p>  |
| 1              | Trialing solution(s)                           | The findings are discussed and compared with students' own experiences and observations. Functional research is conducted on the topic: various experiments and study of different data sources.   |

|   |  |  |
|---|--|--|
| 1 | Assessing solution(s)                          | Based on what they have learned, student groups construct models of the functions of the body.   |
| 1 | Discussing solution(s)                         | The models and reports are presented and discussed.  |
|   | TEACHER REFLECTION                             | Use of the digital tool for formative assessment of competences will continue in the same way during the second learning cycle.  |
| 2 | Defining a real world problem related to SDG's | How different senses work and how they would remain as functional as possible: Ear and Sound, Skin and Touch, Eye and Light.<br>Information reading circuit on sensory function, notes on how different senses work. |
| 2 | Finding solution(s)                            | Different tests are performed for the different senses, as well as notes from observations and conclusions.  |
| 2 | Trialing solution(s)                           | Tests are repeated to confirm the conclusions  |
| 2 | Assessing solution(s)                          | The results are compared. The students make their own presentations / videos about their own sensory activity station.   |
| 2 | Discussing solution(s)                         | The activity stations related to the senses are presented to others. Feedback is received from peer groups about the activity stations via Qridi group evaluation.   |

At the end of learning cycle 1, students were asked to:

1. Post a photo of their the body part related activity station that they build during the project
2. Answer three reflective questions with full sentences:
  - Did all members of the group equally participate in the work (division of labor)?
  - Did you try your best in all the tasks assigned to you?
  - Did you find the most important information about your topic, and were you able to write them clearly in the summary?
3. Give their group a grade for working as a group that day with the scale of beginner / moderate / good / excellent. A brief description of success criteria is provided for each of the four grading options.



|  |  |
|--|--|
| Targeted transversal STEM competencies | <ul style="list-style-type: none"> <li>● Collaboration</li> <li>● Communication</li> </ul>   |
| Digital tool(s) utilized               | Qridi  |
| Formative assessment strategies        | Digital tool (Qridi) utilized for: <ul style="list-style-type: none"> <li>● Self-assessment</li> <li>● Peer assessment</li> <li>● Sharing learning intentions</li> <li>● Clarifying success criteria</li> <li>● Classroom discussions</li> </ul> |
| Artefacts produced during the cycle 1  | Students produce a project report about metals based on their own research.  |
| Artefacts produced during the cycle 2  | Students produce a video showcasing their own metals research. Topics include the ecosystem of metals, corrosion, and recycling.   |

Table 4: Learning cycles 1 and 2 for school B

| Learning cycle | Step   | Explanation  |
|----------------|--|--|
| 1              | Defining a real world problem related to SDG's | <p>What metals are, where they are obtained, where they are needed, and what disadvantages they can have.</p> <p>Each double lesson (2x45min) of the cycle covers all the steps, exploring the properties of metals and considering their relevance to Agenda 2030 goals and the ecosystem, such as climate change, sustainable industry and production, sustainable cities and development, equality, etc.</p> <p>The topics are:<br/>           Investigation of the reduction of copper and the hardening and tempering of steel wire; Investigation of the electrical and thermal conductivity of metals; and Investigation of the reaction rate and nobility of metals.</p> |
| 1              | Finding solution(s)                            | In each lesson, students study the given topic in a small group.   |
| 1              | Trialing solution(s)                           | The solution is tested, chemical experiments are conducted on the topic, and the findings are reported.  |

|   |  |  |
|---|--|--|
| 1 | Assessing solution(s)                          | The experiments are repeated so that solutions can be assessed and conclusions can be drawn.   |
| 1 | Discussing solution(s)                         | Students' own research results are compared with the research results of others, as well as with empirical scientific research, and conclusions are drawn.   |
|   | TEACHER REFLECTION                             | The teacher, in consultation with a colleague, considered reducing the amount of reflective questions that Qridi is used for, for the sake of clarity, but did not end up changing its use. That is, use of the digital tool for formative assessment of competences will continue in the same way during the second learning cycle. |
| 2 | Defining a real world problem related to SDG's | The teacher prepares topics, such as "metal recycling," "corrosion," and so on, from which students set out to conduct research based on what they have learned earlier.   |
| 2 | Finding solution(s)                            | Students search for information from the textbook, other books, and various websites.  |
| 2 | Trialing solution(s)                           | Pupils write their findings on a common platform and build up information on the topic based on this to create a script for a video. They also search for information that would contradict or problematise their own perspectives and consider the matter.  |
| 2 | Assessing solution(s)                          | Students work on a video on the chosen topic. At the same time, they consider whether the video is logical and whether it solves the given problem. In other words, assessing the suitability of the solution was also discussed at this stage.  |
| 2 | Discussing solution(s)                         | The videos produced by each group are presented to other groups. Each video is discussed and receives feedback.  |

At the end of each lesson in learning cycle 2, student groups are guided to discuss their work, and mark their evaluation in Qridi. Teacher will also assess student groups after each lesson, and mark his evaluation with comments in Qridi before the next lesson. On a scale 1-5, the student groups were asked to assess their success in:

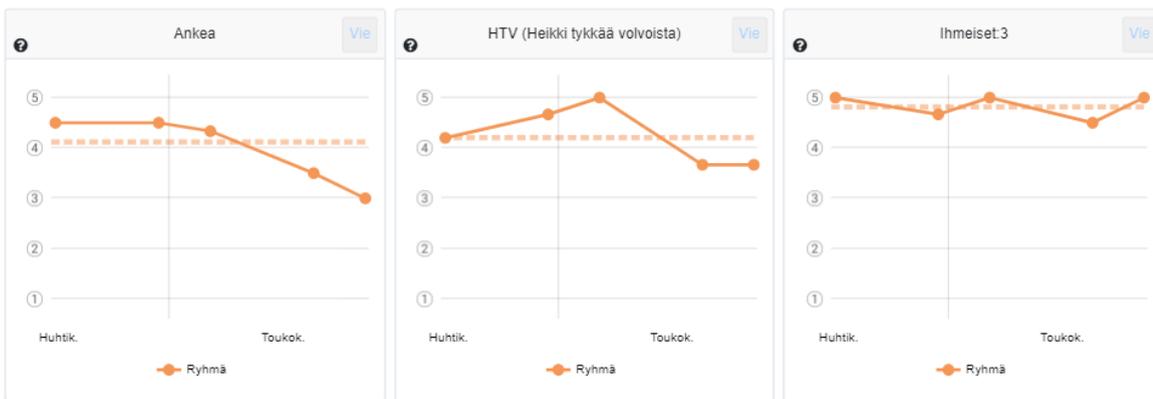
1. communication skills

- a. we listened to each other
- b. we spoke alternately
- c. we discussed at the appropriate volume
- d. everyone's thoughts were taken into account in the work
- e. we negotiated about the solutions
- f. we spoke kindly to each other

## 2. group working skills

- a. we focused on working
- b. we cleaned up after ourselves
- c. we stayed on schedule
- d. we understood the assignment
- e. we accomplished the things we planned to do
- f. we knew how to ask for help
- g. we planned to do the work evenly among the group members

Figure 3: An example of student groups' assessment in Qridi in school B



## 2. METHODOLOGY

### 2.1. DATA COLLECTION

#### 2.1.1. Data collection process

##### 2.1.1.1 Classroom observations

Table 5: observation activities conducted by observer 1

| Date      | Duration        | Learning cycle (1-2) | Step (1-5) | School A / School B |
|-----------|-----------------|----------------------|------------|---------------------|
| 8.4.2021  | 2*45min lessons | 1                    | 1+2        | A                   |
| 15.4.2021 | 2*45min lessons | 1                    | 3+4        | A                   |
| 20.5.2021 | 2*45min lessons | 2                    | 3+4        | A                   |
| 27.5.2021 | 2*45min lessons | 2                    | 3+4+5      | A                   |
| 14.4.2021 | 45min lesson    | 1                    | 1          | B                   |
| 15.4.2021 | 2*45min lessons | 1                    | 1+2        | B                   |
| 22.4.2021 | 2*45min lessons | 1                    | 1+2+3      | B                   |
| 20.5.2021 | 2*45min lessons | 2                    | 3+4        | B                   |
| 27.5.2021 | 2*45min lessons | 2                    | 4+5        | B                   |

Table 6: observation activities conducted by observer 2

| Date      | Duration        | Learning cycle (1-2) | Step (1-5) | School A / School B |
|-----------|-----------------|----------------------|------------|---------------------|
| 8.4.2021  | 2*45min lessons | 1                    | 1+2        | A                   |
| 15.4.2021 | 2*45min lessons | 1                    | 3+4        | A                   |
| 29.4.2021 | 2*45min lessons | 1                    | 3          | A                   |

|           |                 |   |       |   |
|-----------|-----------------|---|-------|---|
| 20.5.2021 | 2*45min lessons | 2 | 3+4   | A |
| 27.5.2021 | 2*45min lessons | 2 | 3+4+5 | A |
| 15.4.2021 | 2*45min lessons | 1 | 1+2   | B |
| 22.4.2021 | 2*45min lessons | 1 | 1+2   | B |
| 6.5.2021  | 2*45min lessons | 2 | 1+2   | B |
| 20.5.2021 | 2*45min lessons | 2 | 3+4+5 | B |
| 27.5.2021 | 2*45min lessons | 2 | 4+5   | B |

### 2.1.1.2 Interviews

Table 7: Teachers interviews

| Interviewee | Date     | Online / f2f |
|-------------|----------|--------------|
| Teacher A   | 1.6.2021 | f2f          |
| Teacher B   | 1.6.2021 | online       |

Table 8: Student interviews

| School A / School B | Number of interviewees | Age of interviewees | Gender of interviewees | Date     | Online / f2f |
|---------------------|------------------------|---------------------|------------------------|----------|--------------|
| A                   | 5                      | 11-12               | 2 girls, 3 boys        | 1.6.2021 | f2f          |
| A                   | 4                      | 11-12               | 3 boys, 1 girl         | 1.6.2021 | f2f          |
| B                   | 2                      | 14-15               | 2 boys                 | 1.6.2021 | f2f          |
| B                   | 2                      | 14-15               | 1 boy, 1 girl          | 1.6.2021 | f2f          |
| B                   | 3                      | 14-15               | 1 boy, 2 girls         | 1.6.2021 | f2f          |
| B                   | 3                      | 14-15               | 3 girls                | 1.6.2021 | f2f          |

### 2.1.1.3 Selection of artefacts

#### 2.1.1.3.1 School A

From school A, a total of 3 artefacts were collected during the first learning cycle. The artefacts consist of the final products produced during the learning cycle ie. constructed models of heart and lung function, together with design plans filled in during the learning cycle (initial designs for the model, equipment needed, research observations, and reflections about what could have been done differently). The artefacts were chosen by teacher A, based on the quality of the artefact, and demonstration of targeted transversal competences during the work process.

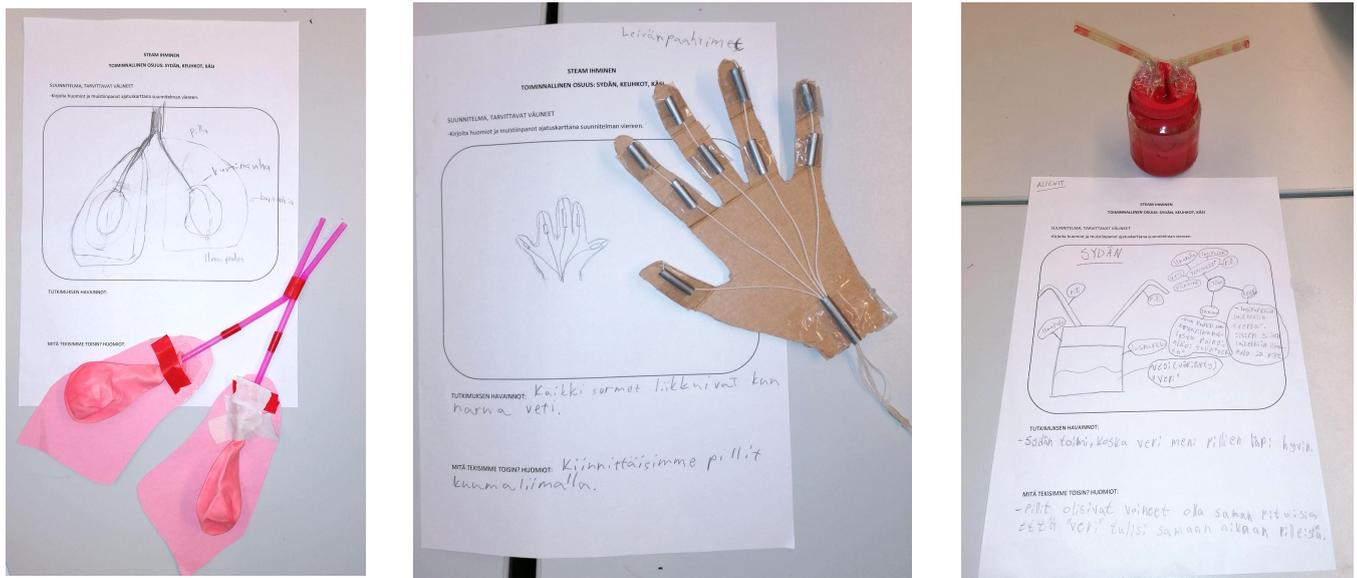
For the second learning cycle, a total of 3 artefacts were collected from school A. The artefacts consist of student videos produced at the end of the second learning cycle. The artefacts were chosen by teacher A.

Table 9: Example of reasoning for selection of artefacts from school A

| Level of artefact | Reasoning for selection of artefacts from learning cycle 1  |
|-------------------|---|
| Good              | <p>The group worked well together and did e.g. discuss the division of labor fluently. They developed their work e.g. staining the water with a suitable color to resemble blood and repairing the attachments after testing the yield several times.</p> <p>The structure of the model worked well. The group developed the structure through testing. The group did a good job with planning, research findings and producing ideas for further development of the model.</p>   |
| Intermediate      | <p>There are some challenges at the beginning of the project, e.g., discussion of division of labor and equal participation of all. However, during the construction of the activity station, all students were involved. They did not use Qridi for assessment, possibly because the students were not in school that day?</p> <p>The activity station is well constructed but based on an already existing idea. When it comes to the planning and research findings, everything mandatory has been completed.</p>  |
| Poor              | <p>The model is not finalised and it is based on an already existing idea. The plan has been made, but research findings and areas for development are missing.</p> <p>The group was not able to share work at the beginning and the construction was pretty much left to just two of the students. In this group, there is one special needs student who does not manage to focus on working together and this is reflected in the division of labor of the whole group.</p> <p>In the end, the work goes smoothly, but the output itself and its finishing are missing even though there would have been time for it. Planning and research findings have been well documented by</p> |

the two active team members. Interestingly, in the self-assessment, the group has assessed its expertise as excellent.

Picture 3: Examples of artefacts from school A, learning cycle 1



### 2.1.1.3.2 School B

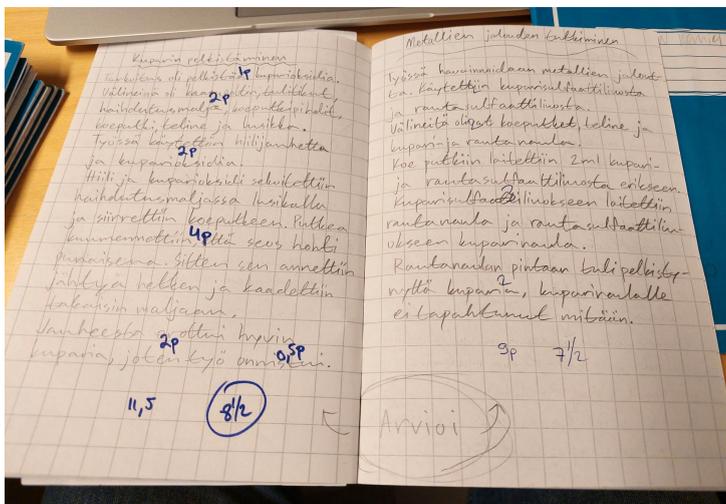
From school B, a total of 3 artefacts were collected during the first learning cycle. The artefacts consist of student work reports produced at the end of the first learning cycle. The work reports present the purpose of the work, equipment needed, chemicals and reagents used, steps to implement the experiment, results, and analyses and comments. The artefacts were chosen by teacher B.

For the second learning cycle, a total of 3 artefacts were collected from school B. The artefacts consist of student videos produced at the end of the second learning cycle. The artefacts were chosen by teacher B, based on the objectives set for producing the video and observations of the group work.

Table 10: Example of reasoning for selection of artefacts from school B

| Level of artefact | Reasoning for selection of artefacts from learning cycle 2  |
|-------------------|---|
| Good              | Both video and teamwork were on a good level. Although group work is difficult to verify because the group worked quietly while co-producing their shared document in OneNote. Students say they have shared responsibilities for searching for information and writing. One student was responsible for preparing the video, and then completing it with the others.                                       |
| Intermediate      | Intermediate video but good teamwork. The division of work was smooth, even, and respectful of their strengths.   |
| Poor              | Poor video and teamwork. One student was responsible for making the video and that student was not in school during the last lesson of the project. Not much discussion about working skills, division of labour, etc. No shared online work since that one student was responsible for producing all of the video material. No self-produced text or material in the video, only photos from the internet. |

Picture 4: Example of an artefact from school B, learning cycle 1



#### 2.1.1.4 Observations in the virtual environment: number and process of implementation

**Table 11: Observation in the digital tool Qridi**

| School A / School B | Learning cycle | Observation activities  |
|---------------------|----------------|---|
| A                   | 1              | The virtual environment, Qridi, was observed two times during the learning cycle. The observer made notes and took screenshots in order to document the observations in the virtual environment.  |
| A                   | 2              | The virtual environment, Qridi, was observed two times. Once during the learning cycle, and once after the learning cycle. The observer made notes and took screenshots in order to document the observations in the virtual environment. |
| B                   | 1              | The virtual environment, Qridi, was observed once during the learning cycle. The observer made notes and took screenshots in order to document the observations in the virtual environment.   |
| B                   | 2              | The virtual environment, Qridi, was observed two times. Once during the learning cycle, and once after the learning cycle. The observer made notes and took screenshots in order to document the observations in the virtual environment. |

#### 2.1.1.5 Electronic questionnaires

The collection of quantitative student questionnaires was implemented electronically by the University of Santiago de Compostela. The analysis of student questionnaire data from all ATS STEM partner countries, was also analysed by the University of Santiago de Compostela. For the collection of quantitative data, the Finnish teachers were provided with a Finnish translation of instructions written by the University of Santiago de Compostela, including examples to clarify challenging concepts contained in the questionnaire.

In Finland, the pilot schools collected the questionnaire data on the implementation of the first lesson of the ATS STEM project, and the final evaluation was completed during the last lesson. Each school determined for itself where in the school year the ATS STEM project will be located. In September 2019, the first pilot schools' implementations started, and in April 2021, the last pilot school commenced their first learning cycles. The first pilot project was completed in December 2019, and the last one in June 2021.

### **2.1.2. Incidents to be taken into account in data collection**

In Finland, formal written lesson plans are not standard. Since the teachers took part in ATS STEM project voluntarily, and were working under time constraints, no formal lesson plans were produced.

The gender ratios of the students taking part in the ATS STEM projects at participating schools are estimates, other than for case study schools where actual numbers were collected by the observers.

From school A, the interview of the student group who produced the intermediate artefact had been planned, but due to the changes in the school's schedule was not able to be conducted.

In collecting classroom observation data, the observers did not count the number of instances of different indicators that express frequency, such as sometimes, often, etc. As such, these are subjective measures of these indicator questions.

The student questionnaire from which quantitative data was collected, contained several concepts unfamiliar to the students, such as hypothesis, systemic thinking, or metacognition, which, despite the presence of additional instructive examples, could possibly affect the quality of student answers.

## **2.2. DATA ANALYSIS**

The data from the different sources was processed transversally, formulating initial responses for each research indicator. The analysis progressed from initially focusing on each indicator question independently, to then analysing all of the indicators of each subcategory collectively, to finally analysing and interpreting each category on the basis of the analyses of the subcategories. Thus the analysis of each of the six dimensions or categories of the ATS-STEM model incorporates the analyses of the subcategories and indicator questions, and these explanations and interpretations are developed based on the different sources of qualitative data collected during the project. The analysis was based on the evidence collected in the study, avoiding the researchers' personal biases and beliefs.

### **2.3. QUALITY GUARANTEE**

The data was extracted from different sources: documentation of the case studies; interviews of students, teachers and the mentor; observations of both classroom activities and the digital assessment platform Qridi; and artefacts produced by the students during the project. For both classroom and digital assessment platform observation, the same observer observed at different times (ie. stability technique, Bisquerra, 1989).

The participants, and their parents in the case of students, were informed of the ATS STEM project and data collection. Consent for the use of the data was granted only for use in the ATS STEM project. Anonymity is ensured by referring to cases and participants with codes. No personal data was collected. The data was stored at TUNI following GDPR (EU 2016/679).

### 3. RESULTS

#### 3.1. RESULTS DESCRIPTION

##### 3.1.1 Digital assessment description

The case study schools utilised Qridi as the digital tool for formative assessment. All students of the case study schools had some prior experience using Qridi, but nevertheless all students received training in how to use Qridi at the beginning of the project.

Table 12: Formative assessment strategies for school A

|   | Sending and/or Displaying  | Analysing and/or Processing  | Interactive environment |
|---|--|--|-------------------------|
| Sharing learning intentions and clarifying success criteria | The learning objectives and success criteria are presented in Qridi. |  |                         |
| Questioning and classroom discussions                       |  |  |                         |
| Giving and using feedback                                   |  | Teacher follows up self-assessment and peer assessment evaluations made in Qridi by initiating face-to-face classroom discussions. |                         |
| Self- and peer-assessment                                   |  | Students perform self-assessment evaluations in Qridi.   |                         |

Table 13: Formative assessment strategies for school B

|   | Sending and/or Displaying   | Analysing and/or Processing  | Interactive environment |
|---|---|--|-------------------------|
| Sharing learning intentions and clarifying success criteria | The learning objectives and success criteria are presented in Qridi, and students review them when doing self-assessment and peer assessment. |  |                         |
| Questioning and classroom discussions                       |   |  |                         |
| Giving and using feedback                                   |   | Teacher follows up self-assessment and peer assessment evaluations made in Qridi, and give their own feedback after each lesson. |                         |
| Self- and peer-assessment                                   |   | Student groups perform self-assessment and peer assessment evaluations in Qridi.   |                         |

### 3.1.2 Formative assessment achievements

#### 3.1.2.1. Integrate STEM content

The learning objectives are presented to the students in a digital format, and are also explained as part of an oral presentation at the beginning of the project in both schools, according to the observers. Teacher A from school A elaborates that at the beginning of the project, “teachers and students looked at the SDGs and formulated learning goals for transversal competencies together”. In the beginning of the lessons, teacher A “reminds students of what they will be assessing at the end of that lesson”. In some lessons, students assess group working skills. This assessment happened in the digital tool Qridi, and “students have to use their own words to describe their assessment”. Teacher B from school B states that “the transversal competency goals are presented with a powerpoint presentation at the beginning of the project, and they are presented in Qridi [the digital assessment tool] so students see them at every lesson.” The learning objective or goal in school B “was about group working skills”, and teacher B comments that “having to assess their skills each lesson helped the students” to work in groups “a bit better than before”. Teacher B had “wanted to discuss with each group after each lesson to give detailed feedback to help them but did not have time”, but believes it “would be beneficial” and wants to do more of this in the future.

The students of school A say that understanding and remembering the learning objectives is potentially challenging. Both the poor and good artifact producing groups from learning cycle one agree that it “can be difficult” for the students to “understand the learning objectives, if they are only presented by the teacher in the beginning of the project” The good artifact producing group think “It helps” when teacher A reminds them during the learning cycles about “what has been done, what happens now, and what is yet to come”. When asked in the interview, the poor artifact producing student group remembers which skill was targeted in the project, but are not able to specify the skill in more detail than the topic. They also recall teacher A reminding them of “a blue objective table that was presented at the beginning of the project and where it is on the classroom wall”, which helps the students remember the goals. Teacher A confirms that in explaining the learning cycle process to the students, “the phases are visualised on the classroom wall”, and that “it is explained often”. Teacher A adds that “some students are able to use the discussed learning goals to guide their actions and even guide their group by reminding them that some things need to be done especially well because those things are being assessed”. Yet, “some students are not able to use the discussed learning goals to guide their actions, even though they know what is being assessed”, and there is “large variation in this”.

The observers note that in school B the learning goals of group work and independent work were reviewed together. The groups were mainly able to describe the competencies in interaction skills, some very well. The students were also “provided with the opportunity to adjust the learning goals for teamwork”, according to the observers. It “seemed that

the learning goals were familiar to the students". They "agreed that the goals of group work were suitable", and "the students were able to describe the goals with different examples", but they "did not have any further questions". The students of school B had somewhat different interpretations of how the learning objectives were presented. The poor artifact producing student group from learning cycle two said that they were "informed about the objectives at the beginning of the project", and they were "re-introduced during the project", "not in every lesson, but at times". This group of students said they "did not think about the targeted competences [of group working skills and communication] when [they] worked". At the end of the project, these students are not able to verbalise what skills they have learned during the project when asked to describe with their own words, but when the interviewer reminds them, then they remember and agree that they "have learned those skills". Students from the intermediate artifact producing group said that teacher B "did not inform [them] about the objectives at the beginning of the project", but "only showed slides that presented the group work skills", which, however, "did not make clear what [they] were expected to learn". The good artifact producing student group "do not quite remember" what was mentioned at the beginning of the project, but say that the "questions in Qridi were clear and made [them] think about the group work skills." Even though the students discuss the learning process of the skills really well during the interview, at the end, when asked, they are not able to name skills they learned during the project. Teacher B believes that it is possible to guide students to understand that they need to follow a certain learning process, but "analysing and discussing the learning and working process is difficult for most students". Teacher B "encourages students" but would want them to "learn to be more independent and skillful in improving their product by assessing and discussing it within the student group", so that the students would "see the value of that process", without the teacher "having to pressure them to do it".

The different artifact producing students groups of school B feel that they understand the assessment. The poor artifact producing group from learning cycle two say "it was clear [to them] what the main topics of assessment are." They do not remember if they had the possibility to affect the success criteria, but teacher B "gave criteria for assessment", and they "remember that the criteria was discussed". They believe they "understood the basis of assessment of the project", and they "think [the assessment] works well." The intermediate artifact producing group also think that the assessment of the project was "clear", but in elaborating on this it is evident that they mean that they understand that "30% of grade is based on classroom work, 30% on the work report", etc, and not so much the assessment criteria itself. Teacher B thinks that the students are very goal-oriented in their work, saying that the students themselves are very "focused on grades", which can be leveraged to get them to "focus on group work skills". For this reason, teacher B plans to have 20 percent of the students' grades based on group working skills in the future.

The teachers from both schools taught the two learning cycles alone to their respective classes. In both schools, there were no co-teachers for the observed lessons. In school A, "apparently several teachers took part in the planning of the learning cycle and project", although the two learning cycles were taught by teacher A, a classroom teacher, alone.

Similarly, in school B “co-planning of teachers has taken place earlier”, although “there is no co-teaching” in the observed lessons. which were taught by teacher B, a chemistry teacher, alone. During classes, teacher B “refers a couple of times to the contribution of other teachers” of biology and geography in the learning cycles, and “the connection between biology and geography and the topic of metals” is mentioned. The observers note that teacher B “works alone in his own subject”, but other teachers also have Agenda 2030 social development goals as a theme, and “they have planned the theme together in the past, but each takes it from within their own subject”. In the first lesson of learning cycle one, teacher B “refers to a common theme with other teachers” in other subjects. Teacher B confirms that there is a “shared topic” of Agenda 2030 for “all eighth grade students”, but elaborates that “teachers implement it in their own teaching”. Teacher B expresses interest in collaborating with colleagues to “develop assessment together”, and in the future to “implement shared assessment of interdisciplinary learning modules”.

In school A, the theoretical and practical knowledge of the subjects Environmental Studies, Finnish language, Mathematics, Social Studies, and Visual Arts are all necessary, and are combined in the two learning cycles of the project. Teacher A comments that because “the topic was the human body”, “it was easy to connect theoretical knowledge and everyday life”. In school B, chemistry was the only subject from which theoretical and practical knowledge was necessary for the learning cycles, although teacher B does refer to biology and geography in the learning cycles, and “the connection between biology and geography and the topic of metals” is mentioned. Teacher B explains that “students first studied theory, then did practical experiments and tests, then produced a video in groups. When producing the video, students quickly found the theoretical information again from books, online, etc.” Teacher B “would like to go deeper into the theory in the future, thinking that it “was too much just searching for information” of the topic and too little time to deepen knowledge”. The student group that produced the poor artifact for learning cycle two says that the students “are able to use theory that [they] learned from the textbook when doing a research project, to some extent”. The intermediate and good artifact producing groups concurred.

In the course of the learning cycles, explanation of the relationships between the disciplines, and the use of the knowledge of different disciplines varies. The observers note in school A, that at the beginning of the first lesson, “the different subjects combined in the learning cycle are explained”. Teacher A elaborates that “At the beginning of the project, students are informed about the learning goals of each subject involved, and how their progress will be assessed.” Teacher A gives an example that “if the students need to write a report about a topic they have been researching in science, they will also know which things about the use of Finnish language will also be assessed in the report.” In school B, the observers note that there is “hardly ever” any explanation of the relationships between different disciplines, “since there is only one subject”, but regarding the use of knowledge of the different disciplines “reference is made to writing a research report” using Finnish language and writing or “compiling a video” using information technology or computer skills.

The students themselves are capable of discussing and reflecting on how the interdisciplinary knowledge and theory can be used to help to solve the tasks in the learning cycles. From school A, the poor artifact producing group from learning cycle one says in their interview that “It helped to build an activity station, when [they] had first studied the topic from books. [They] feel like [they] could not have done the activity stations if [they] didn’t first study the topic.” Similarly, in school B, the poor artifact producing student group from learning cycle two believes that the students “are able to use theory that [they] learned from the textbook when doing a research project, to some extent”. Importantly, the intermediate and good artifact producing groups each provided similar answers to this question. The observers note that the students in school A can reflect on how the interdisciplinary knowledge helps them in solving the tasks. In contrast, the observers state that students in school B “hardly ever” or “never” discuss and reflect on how the interdisciplinary knowledge helps them in solving the tasks, other than “one of the students stat[ing] that it is easy to make a video, because it has already been learned in the 7th grade in ICT”.

The evaluation criteria were developed primarily by an individual teacher in both schools. Teacher A and a colleague “developed the evaluation criteria together earlier”, but teacher A “specified them at the beginning of the learning cycles”. The observers note that teacher A encourages the students to help “refine the learning goal”, but that “the main structure was agreed between the teachers and taken to the students.” The observers state that the evaluation criteria in school B were developed independently by teacher B, but the criteria were “looked through together”, and the “number of goals were discussed” with colleagues from school B.

In both schools, the students’ work is evaluated by a single teacher, so there is no co-evaluation by a team of STEM teachers. The observers note that both teacher A and teacher B evaluated the work of their students alone.

### 3.1.2.2. Reflect STEM learning design principles

Observations show that the teachers have tried to have the lessons address real world problems and challenges. Both schools were positively assessed by the observers as encouraging a vision of how to bring this problem to real life, linking research to the real world, and most of the time encouraging the building of solutions using new knowledge. The observers' assessment of raising problems connected to the real world was "Yes, but only in some activities/tasks" for both schools. The observers noted that in school B, although the chosen topic of metals was rather challenging and theoretical, the teacher still "directs the assignments so that the practical connection appears in almost all the lessons observed." According to the observers, this was achieved through "descriptions of everyday situations, such as hardening of metals or why it is needed, and focusing on the metal mining industry in Finland". The students of school B expressed some doubt about the connection between the topic and the real world. The poor artifact producing group felt that "the topic was not related to everyday life" even though "sometimes, the teacher tries to help connect the topic to everyday life, but not often". The intermediate artifact producing group said that "the topic was related to everyday life if you happen to end up in a certain profession". The good artifact producing group were somewhat more positive, stating that "the project was related to the real world, but not to [their] own lives". School A chose a more tangible topic about the human body, which teacher A believed "was easy to connect with everyday life". This was confirmed by the observers, who noted that students were encouraged to "look for real life approaches, for example, by testing muscle tension, wondering what kind of food is healthy and what is not healthy, and why so." Students conducted research linked to real world problems by, for example "looking at which foods in school lunch contain fat and which do not".

Input from all team members was "always" encouraged for both schools. According to the observers, this was essential in the case of school A, as the "non-fiction book club" group format for sharing ideas required all team members to perform a designated role, so that "everyone has their own important task." The observers similarly note that within the groups it is possible to divide the work between participants according to each member's strengths. Teacher A agrees, stating that "each student has had a certain role in the group", and that "roles are divided between the students themselves based on what they think is the best solution." The observers note that in school A, "some students are really active and in a few groups the responsibility is nicely and evenly distributed, but there are also those groups where all the work and responsibility falls to a few students". As part of their group assessment, students of school A are asked to answer the reflective question, "Did all members of the group equally participate in the work (division of labor)?" Teacher A does "remind students several times during the project about trying to learn versatile things, not only to always want to do something they are already good at in the group", even though the division of labour within the class groups is freely decided by the students themselves. Teacher A's analysis of the artifacts produced in school A note that the level of collaboration in group work ranged from "working well together" and "discussing the division of labor fluently", to having "some challenges at the beginning of the project discussing the division of labor and equal

participation of all”, while still having all students involved in the construction of the activity station, through to being unable to share work at the beginning so that “construction was left to only two students”. The observers further report that “group work seemed very smooth”, “the division of labor worked well” overall, and there are “no visible challenges in division of labor or joint decision-making”. Although the teacher does intervene, for example, if “any student does not participate in the shared work”, this is “more about a lack of focus or concentration”, than any student being excluded or feeling left out because they don’t get their way. Although teamwork was encouraged in school B, the balance of how that was distributed could be more uneven according to the observers. “Many groups work really quietly and have surprisingly little discussion, especially when making a video, yet they all share their work and do their part. In a couple of groups, work varied and there was more burden placed on the shoulders of just one or two students. In another of the groups, on the other hand, although one member was more active, the others trusted him, and all actively did their part.”

In school A, the observers note, “some students are really active” and “in a few groups the responsibility is nicely and evenly distributed”, but in some groups “all the work and responsibility falls to a few students”. In school B, it is also observed that in one of the groups, “two students seem to make the effort for the project, and the other two students often spend time chatting and doing other things than what they should”, and “this clearly shows in their [self-assessment] answers about sticking with the schedule”. According to the observers, these students in some sense took responsibility for their failure to contribute to the assigned task in their self-assessment answers. Thus, it is notable that in school B the group assessment evaluation criteria for communication and group working skills included to what extent: “everyone’s thoughts were taken into account in the work”, and “we stayed on schedule”. The observers note that in school B, “groups discuss relatively little during the observations”, mostly related to practical issues such as the division of labour. Discussion occurs mostly during small group work, during which the students “may even give direct feedback to each other”. Many of the student groups “work really quietly and have surprisingly little discussion”, yet they “all share their work and do their part”. However, in a couple of the groups the work was more unequally distributed, and there was “more burden on the shoulders of a couple of students or just one student”. In another student group, one member “was more active” and “was trusted by the others”, even though “all actively did their part”. Teacher B’s analysis of the artifacts produced in school B note that the level of teamwork in group work ranged from “good teamwork” in which “division of work was fluent, even, and respectful of their strengths”, down to poor teamwork in which “one student was responsible for making the video” with “not much discussion about working skills” or division of labour. Indeed, the group assessment criteria for communication and group working skills in school B included to what extent: “we planned to do the work evenly among the group members”. The poor artifact producing group from school B say that teacher B “got everyone working and encouraged them to be active” in the project, and that “the task was broad enough for all students to take part in it”. The intermediate artifact producing group agrees, even though they think “that is the teacher’s job”. The good artifact producing group adds that teacher B “helps group work if needed, but

if the group works well [the students] get to work very independently". In the student interviews the intermediate and good artifact producing groups each elaborate that the division of labour in their respective groups was overseen by one member who was elected as their group's leader.

Encouraging the use of technological or engineering design was somewhat mixed, being assessed as "only in some activities/tasks" for both schools. The observers noted that technological design was used in school B only in some activities, in "planning, producing and editing videos to teach others and to show what has been learned", and in school A students were also offered the opportunity to make videos, even though the use of technological tools were not actually the focus in this learning cycle. However, the observers also noted very strong engineering design principles in the construction of human body parts in school A, writing that "in the lesson, students come up with ideas and construct a different body organ (heart, lungs or hand) from different materials. In this regard, they complete a construction plan outlining the operation of the body part. In this plan, they record what works in the structure and what could have made the structure work even more."

According to the observers, both schools promote cooperative learning primarily through group discussion and group decision-making. School B was assessed as promoting cooperation through planning together, holding meetings for follow-up, and discussion and making decisions, and "depending on the group, all of these methods are used, but mostly discussion and deciding things together", as the students do not spend much time planning more diverse work. The observers note that students from school B could be seen helping each other by, for example "advising where to find the right research tool", or "where to look for the right page in the book". The students' evaluation of the value of group work in aiding shared student learning in school B was varied. The poor artifact producing student group believed that teacher B "encourages students to teach other students what they know", and that they "like working in groups". However, the intermediate artifact producing group was somewhat more circumspect, believing that "group work skills are important for the future work life" but "working in groups does not necessarily help [them] to learn better in school". In their interview, the good artifact producing student group discussed a situation in which the group leader had been thinking she was "too dominant" in explaining tasks and dividing the work, but when she asked her group about this, the other students are supportive of her leadership style and said "it is good that she takes the lead". School A was similarly assessed as primarily promoting cooperation through discussion and group decision-making. "The student groups discuss which of the textbook issues are most important, what kind of questions to figure out from the chapter, what kind of picture is drawn from it, and the like. Whatever is learned in the table groups is then shared with the other groups." As the observers note in reference to school A, it "depends a lot on the student and the group, but generally the groups work well together and help others". Teacher A also believes that "some students are able to use the discussed learning goals to guide their actions and even guide their group by reminding them that some things need to be done especially well because those things are being assessed", even though "there is high variance in this." The

observers similarly note that students have a lot of discussion about what is to be done and how to do it, but here too there are big differences between groups.” Indeed, this is built into the group assessment rubric for school A. Students give their group a grade for working as a group on a four point scale: beginner, moderate, good, and excellent. The rubric for “good” includes “I can often help others” although “sometimes I need help explaining this in my own words”, and the rubric for excellent includes “I can explain this in my own words” and “if needed, I can teach this to others.” So, students supporting each other, explaining to each other how to do the task, and helping each other to keep up is inherent to the assessment and self-assessment of group work. Teacher A thinks “it is important that the students give each other peer feedback” as this is another way that they “can learn from each other”. The student interviews from school A confirm this. Students from the good artifact producing group say that peer feedback “mostly helps them”, and students from both the good and poor artifact producing groups say that they “enjoy group work”, and think that “learning in [and from student] groups can happen”. Further, the observers of school A reported that “group work seemed very smooth. The division of labor worked well”. Teacher A says that learning groups are decided by the teacher and each student has a particular role in the group. At this beginning, however, “they only have the group leader and other roles are divided between the students themselves based on what they think is the best solution.” They have practised this for several school terms. Teacher A thinks that “the most important tool is discussion”.

The observers assess the students of both schools as having autonomy in decision-making about its own learning process most of the time. The observers note about school A that the student group also “has the opportunity to influence the things to be assessed”. Teacher A clarifies that “at the beginning of the project, teachers and students looked at the SDGs and formulated learning goals for transversal competencies together.” In school A, the observers note, the student groups “are quite free to think about the most important things, but the teacher directs them to focus particularly on important topics”. Likewise the division of labour within the class groups is freely decided by the students themselves, but teacher A does “remind students several times during the project about trying to learn versatile things, not only to always want to do something they are already good at in the group.” The observers similarly note that within the groups it is possible to divide the work between participants according to each member's strengths.

Observers report that for both schools the classroom environment is safe, and that students are able to express themselves without fear of making mistakes. This is reported as being the case most of the time for school B and between always and most of the time for school A. Both of the student groups interviewed from school A largely agree, saying that “it feel it is ok to make mistakes in the classroom”, and “sometimes [they] might feel a bit embarrassed but there is no bullying or such because of making mistakes”. The observers note that school B has a “good atmosphere”, but that “a few shy or quieter students work quietly throughout the class”. School A, at least from the observers’ outsider perspective, “seems to have a really safe classroom environment, but it's hard to say anything about an individual student's experience”. Teacher B notes that “many kids are focused more on grades than learning”, and on that basis can

be somewhat risk averse, fearing that making mistakes can lead to lower grades. So teacher B is now trying to turn this around "by having fewer tests and making group assignments or reporting assignments as the basis of grading". Teacher B also commented that they "like to keep students working in the same groups for a longer time" so that "95% of students are able to work well in groups and feel good". Teacher A thinks that the classroom atmosphere is mostly safe and good. Mistakes are not something students want to emphasize, especially when presenting ideas in front of the class, but the teacher thinks that is natural for people. However, "when students are building something with their hands, they are not as afraid of mistakes". He hopes that the same courage would be present in verbal discussions as well. The poor artifact producing student group feels that "it is ok to be yourself and make mistakes in the classroom" and the intermediate artifact producing group agrees, but thinks that "it is also up to yourself. Everyone has their own limits of when they think a place is safe".

The teachers in both schools try to conceptualise failure and error as indispensable to the learning process. The observers note that "yes" they are indispensable in both schools, and "always" so for school A, as "in this form of learning, i.e. in the "non-fiction book club", it is more difficult to fail directly, but ideas are constantly being worked on to become more functional". Teacher A also "reminds students several times during the project about trying to learn versatile things, not only to always want to draw and doing something they are already good at." Teacher B thinks that students are getting better at tolerating mistakes. In the final report, more points are given if the students also reflect on their mistakes and don't just say "it all went well". Teacher B also presents "good model answers of previous students who have reflected on the mistakes well" and also that explain "what was learned because of the mistakes". Teacher B also understands that this is challenging for the students, noting that "the best students can do it, but the rest of them cannot do it just yet." Teacher B would like to instruct this even better in the future.

The observers state about both schools that yes, the teachers encourage critical thinking, and yes, the classroom is a space for experimentation and creativity, in which students are encouraged to experiment and seek their own answers. Similarly, the observers affirm that the activities in both schools motivate students to explore new knowledge for themselves. For school A, the students "can search for information in the activity station from anywhere, not just from the book", so students' exploration of new knowledge is encouraged. In the analysis of the artifacts produced in learning cycle one, teacher A highlights testing and experimentation as something that sets the good artifact apart from those produced by the other student groups, noting that the "group developed the structure through testing", produced "ideas for further development of the model", and they also "developed their work" by experimenting with dyeing the water a "suitable color to resemble blood", and also made repairs "after testing the yield several times". For school B, most of the time exploring new knowledge is encouraged, and "it would be possible to be really active" in seeking out new information for the assignments and "one group is especially proactive", but some students don't necessarily make the most of the opportunity provided.

Similarly, the observers assessed that students are mostly provided with activities and problems that are open enough for them to explore their own insights. The observers noted that for school A, yes, the assignment of the activity station is quite open, and :leaves a lot of space for the students' own ideation and planning", even though teacher A "is always close by to support and guide". Teacher A affirmed this, saying that he thinks that "the end product for cycle two was very open. Students were only given the topic and how they wanted to build the activity station was then totally up to them. However, the students have already been trained step-by-step to be able to do this type of activity before." School B was assessed by observers as allowing students to explore their own insights, "only in some activities and tasks". However, the poor artifact producing group from school B felt that "the task was broad enough for all students to take part in it". Although the students are encouraged to explore, Teacher B "would want the students to learn to be more independent and skillful in improving their product by assessing and discussing it within the student group." In this way the students would see the value of that process, without the teacher having to pressure them to continue to explore and improve. Teacher B believes that "teaching this needs to be more systematic."

Observers noted that both schools encouraged the student groups to share the results of their work. Students from school A "independently trialled their activity stations by presenting them to the other students" and also shared them with younger students from the school. The students from school B, although only in some activities and tasks, did also share the videos they created with other students.

### **3.1.2.3. Facilitate feedback**

Individual teachers are responsible for giving feedback to the students for the different tasks they undertake in both schools. In any case, the STEM project learning cycles were primarily implemented by individual teachers in both schools.

There was always consistent feedback from teacher A on learning achievements of students throughout both learning cycles in school A, according to the observers. Teacher A "walks around the class and presents the students with development ideas, and praises when something is working in the groups", so "the students receive feedback on their work throughout the learning process". In school B, as the observers note, feedback tended to be primarily summative, and only in some tasks for the first learning cycle, but formative feedback increased for the second learning cycle. For the first cycle, there is some formative feedback provided "in group work and research tasks, but primarily "research reports are collected at the end of the learning cycle". Whereas in the second learning cycle, teacher B "gives the students feedback on the research based on the plan in writing, but also orally while the students are doing the research." The teacher and students "go through the group work assessment surveys together, and evaluate the working

methods and the accuracy of the results orally” during each lesson. Teacher B agrees that “there should have been better group work feedback after the first learning cycle”. The students of school B think that, as the poor artifact producing group said in their interview, teacher B “gave feedback throughout the learning project”.

The observers note that feedback is constructive, providing evidence to support and strategies to improve the work of students most of the time for school B and always for school A. Teacher A “asks guiding questions to help with this”. Similarly, in school B teacher B provides task-centred feedback in a clear and timely manner most of the time and in school A teacher A always does so, according to the observations. In school A, the observers state that there is “continuous feedback in the course of working” and “students also of their own initiative seek open advice”. In school B, the comparative lack of formative feedback in learning cycle one may have limited the extent to which this feedback could have been used to improve student learning for learning cycle two. Indeed, as the intermediate artifact producing student group from school B explains in their interview, “the work report [from learning cycle one] is submitted at the end of the [learning cycle], and gets graded”, but students “do not have the opportunity then to improve it based on the feedback and submit again”. A more formative, rather than just summative assessment of the work reports, giving students the opportunity to practise and resubmit, would help students to introduce changes and improvements in their learning.

Teacher A provides extensive formative feedback, and students “sought information, for example, for their activity stations and are motivated to try what they have learned”. In support of this teacher A says in the interview that “it is important to have enough time to listen to what the students are talking about in the groups, and to then take part in the discussion, to ask for rationalisations and challenge and encourage them.” According to the observers, teacher B is somewhat less consistent in providing formative feedback, especially for learning cycle one, but this may have been partly due to the topic of the learning cycle not being especially conducive to formative feedback or assessment.

The feedback provided gives the students information on their progress towards the set objectives most of the time for school A and sometimes for school B, according to the observations. For school B, feedback provides updates on progress towards the goals of group work. The transversal competences to be assessed in group work and independent work are reviewed together, and the groups are “mainly able to describe the competences in interaction skills, some very well”. Teacher B tries to reinforce this, according to the observers, by “commenting on cooperation”. Teacher B says that the students themselves being very “focused on grades” can be leveraged to get them to “focus on group work skills”.

The students are given many opportunities in school A, and some opportunities in school B, to plan, consider, and review the tasks and the processes followed to solve them. The observers note that In school A, the activity stations

provide “many opportunities to create your own solutions” and “are given plenty of space to generate ideas and solutions under the guidance of the teacher”. The good artifact producing student group from school A agrees, and elaborates that they come up with their ideas for the project by “discussing with classmates and searching online”. In school B, there is limited time and space for considered problem solving during the first learning cycle, which is “really concise, so the subject goals take priority over the transversal competences”. However, in the second round, “there is more time and space for the students’ own solutions”, according to the observations. The poor artifact producing student group from school B say that “the task was broad enough for all students to take part in it”. The good artifact producing student group concurs, saying that “if the group works well [they] get to work very independently”.

The observers note that feedback is constructive, providing evidence to support and strategies to improve the work of students most of the time for school B and always for school A. Teacher A “asks guiding questions to help with this”. The student groups from school B all agree that, as the intermediate artifact producing group phrased it, “feedback helps students to learn better”. The poor artifact producing group further clarifies that the feedback “helps students to change the way they work at that moment but not about how they learn in general.”

Similarly, teachers provide task-centred feedback in a clear and timely manner most of the time for school B and always for school A. The observers state that there is “continuous feedback in the course of working” and “students also of their own initiative seek open advice”. In school B, the comparative lack of formative feedback in learning cycle one may have limited the extent to which this feedback could have been used to improve student learning for learning cycle two.

Teacher A provides extensive formative feedback, and students “sought information, for example, for their activity stations and are motivated to try what they have learned”. In support of this teacher A says that “it is important to have enough time to listen to what the students are talking about in the groups, and to then take part in the discussion, to ask for rationalisations and challenge and encourage them.” According to the observers, teacher B is somewhat less consistent in providing formative feedback, especially for learning cycle one, but this may have been partly due to the topic of the learning cycle not being especially conducive to formative feedback or assessment.

The teachers from both schools provide the students with the assessment criteria for evaluating their own work, according to the observations. In school B, the criteria provided to the students are “for both summative and formative assessment”, and the summative assessment tables are “already familiar” to the students, who are “quite good at assessing their competence, especially in the subject”.

The observers relate that the teachers from both schools sometimes convey the assessments to the students so that they can be used to improve learning autonomy. In school B, teacher B provides oral feedback to the students in most

lessons, but the observers note that there isn't equivalent feedback provided in Qridi or elsewhere. The observers further highlight that this learning autonomy in schools is in general quite "natural in Finland", that "students are allowed to do a lot independently". For example, teacher B encourages students "to find things out for themselves or figure out a task based on their previous experience" in this case in the subject of chemistry. Teacher B, from the teacher interview, "would want students to learn to be more independent and skillful in improving their product by assessing and discussing it within the student group", so that the "students would see the value of that process", without teacher B "having to pressure them to do it".

The feedback from teachers in both schools proposes strategies to improve learning most of the time or always. Teacher A thinks that "the best way to give feedback is direct verbal feedback", and that the "discussion helps the students to improve their performance". The observers note that teacher A "circulates between the groups very naturally, commenting on the group work" and "asking guiding questions". This ongoing feedback and support "helps to make successes visible so that students are able to capitalise on them" to improve their work and learning overall. Teacher B also asks how student groups are doing and "whether everything is fine if the work does not materialise", but does not always "make comments or question processes to help students develop". Even so, the observers note that for teacher B, "oral feedback usually concerns improving the work or research task", and "this happens in all lessons", especially in the first learning cycle. The observers note that teacher B for example suggests information for "deepening the content of the videos" and in other "practical tasks". Teacher B "consistently instructs students to see how the work could be developed", using "introductory questions when conducting research", so that students can evaluate for example "how to ensure that the timing starts for each metal at the same time?" The student groups from school B all agree that the "feedback helps students to learn better". The poor artifact producing group further clarified that the teacher "gave feedback throughout the learning project", and the feedback "helps students to change the way they work at that moment but not about how they learn in general."

According to the observers, students are always encouraged in school A and sometimes encouraged in school B to be actively involved in assessment. In school A, students complete a self-assessment on Qridi at the end of almost every lesson during the learning cycles, so the subjects of the assessment actually progress together all the time. The observers note that "In this way, students' self-assessment is also continuous and timely, so that the things to be assessed are remembered more clearly after the lesson than, for example, if the assessments were only at the end of the learning cycle". Similarly, in school B, student groups are guided to discuss their work after each lesson, and mark their evaluation in Qridi. Still, teacher B believes that success criteria for group work "should be improved and communicated better to the students", and thus intends to "include students in designing the success criteria". As the students are "so focused on grades", the intention is to leverage this to the students to focus on group work skills.

According to the observers, the students of both schools identify most of the time what they have learned and what learning they should have accomplished. In school B, the observers note that “students are able to assess learning, especially when looking at the assessment of group work”, and “it feels like they have been really critical of the whole group”. Similarly, in school A, the observers note that when looking at the groups’ self-assessment in the digital tool Qridi, “the groups had rather critically evaluated their own work.” Teacher A thinks “the students are good at self-assessment”, and can identify how they are learning and progressing. The teacher thinks that “providing the students with assessment criteria helps with this a lot”, and that “this kind of skill [self-assessment] will not develop just by age but needs to be constantly supported in teaching.” Teacher A says that the students can produce full sentences about what was easy and difficult even though they are only in grade 5. In contrast, the observers note that these answers “could have been more justified and do not always come in full sentences like had been asked”.

Similarly the students attribute learning results to their own efforts and responsibility, and recognise what can improve student learning. For school B, the observers note that there is “great dispersion between the students, but to some extent for everyone, yes.” Indeed, in School B, the observers found it surprising that three of the four groups marked in their Qridi evaluations that they felt they “know how to ask for help”, even though they were “working very quietly and independently”, and “it seemed like only one of the groups actually asked for help and discussed their solutions during each lesson”. Evidently, only one of these student groups needed to ask for help and the other two did not actually require help. However, the fourth group “did not ask for a lot of help, but apparently thought that they should have”. Similarly, even within groups, student self-assessment responses match with the observations. The observers note that in one of the groups, “two students seem to make the effort for the project, and the other two students often spend time chatting and doing other things than what they should”, and “this clearly shows in their answers about sticking with the schedule”. Apparently these students are capable of reflecting on what they can and should do to improve their own learning, and can attribute the learning results to their own efforts, even when they fall short of their own aspirations for the research assignments”. The student group that produced the poor artifact in school A said in their group interview that “when teachers discuss with them during the working process, it helps students to improve their work.” Both the poor and excellent artifact producing groups from school A comment in their interviews that they do get constructive feedback from teacher A, but also especially from students from their own working groups.

#### **3.1.2.4. Facilitate peer-assessment**

In both observation schools, the students work together in groups throughout both learning cycles to produce the collected artifacts. Students from school A worked together in small groups in learning cycle one to design plans and construct activity station models of heart and lung function and in learning cycle two to construct activity stations related to the senses, and to produce videos showcasing these activity stations. As noted in teacher A’s analysis of the

artifacts produced, the level of collaboration in group work ranged from “working well together” and “discussing the division of labor fluently”, to having “some challenges at the beginning of the project discussing the division of labor and equal participation of all”, but having all students involved in the construction of the activity station, through to being unable to share work at the beginning so that “construction was left to only two students”. The observers further report that “group work seemed very smooth”, and “the division of labor worked well” overall. Students from school B worked together in small groups in learning cycle one to produce project reports about metals based on their own research and in learning cycle two to produce videos showcasing their own metals research on topics including the ecosystem of metals, corrosion, and recycling. As noted in teacher B’s analysis of the artifacts produced, the level of teamwork in group work ranged from “good teamwork” in which “division of work was fluent, even, and respectful of their strengths”, down to poor teamwork in which “one student was responsible for making the video” with “not much discussion about working skills” or division of labour.

Students in both schools work together in the evaluation process, according to the observations. The observers note that the students from school B, within their own groups, discuss the assessment process somewhat, “especially in the early stages” of the learning cycles. However, “the assessment changes into more individual work as the course progresses”. In school A, teacher A thinks it is “important that the students give each other peer feedback”, and that they “can learn from each other”. Furthermore, teacher A thinks that “the students work harder when they know a peer-assessment is coming up”, so “for example, when the students had built their activity stations, [teacher A] asked other students to test the activity stations too before inviting students from outside of their class to use the activities”. Students from school A say they receive feedback “especially from students in [their] own group”. This is the case for both the excellent and poor artifact groups. Students from the excellent artifact group say that when giving peer-feedback as a group to another group, “some groups had some disagreements but reached an agreement”, but they think “giving peer-feedback is nice and easy”, and that “peer-feedback mostly helps them”.

The teachers of both schools provide the assessment criteria for peer assessment and rubrics to the students from the beginning of the learning cycles to facilitate the evaluation of the work of their peers, according to the observers. In school A, at the end of each lesson, students give their group a grade in the digital assessment tool Qridi for working as a group on a four step scale from beginner to excellent. In school B, the observers note, the assessment criteria are displayed in the digital assessment tool Qridi “at all times”. At the start of each learning cycle the objectives and assessment criteria are reviewed, to facilitate student evaluation of their own work. However, the assessment criteria presented to the students in school B may contain only 1-5 scales rather than detailed descriptions or examples of what actually is required to fulfil each of the 5 levels of the criteria. Indeed, teacher B thinks that “the success criteria for group work should be improved and communicated better to the students”. Presenting the students with clearer assessment criteria that make the learning goals understandable would also be a helpful method of teaching the

students the transversal competence in question. Teacher B already does something similar in providing clear examples when presenting “good model answers of previous students who have reflected on the mistakes well” in their own final reports.

Students from both schools regularly provide feedback on the learning process to their peers. Students in school A regularly assess their own groups’ performance in teamwork on the aforementioned beginner to excellent scale. The observers note that at the end of the lesson, students record a group self-assessment in Qridi, which focuses on “how the group divided the work and how successful it felt in retrieving, recording, and constructing the information”, as well as giving themselves a grade. In addition, the observers note that students from school A also “peer-evaluate each other based on the constructed body, the plan, and the information package compiled from it.” Students set up the activity stations on the tables of their groups, and one of each group remains their activity station’s presenter. The other students rotate from station to station and peer review the structures of the other student groups. In doing so, the students “make use of familiar color codes” that are commonly used in their classroom for assessment, and give the other groups a color tag corresponding to a certain grade for each point” with a good plan given a blue tag or a moderate structure given an orange tag. Teacher A elaborates, saying that at the end of some lessons, the students “used colour coded tables with 4 steps for peer-assessment for the three things they gave peer-feedback about” (elaborated in 1.3.2.1 Case study school A). Teacher A thinks it is “important that the students give each other peer feedback”, and that “they can learn from each other”. Teacher A believes that the students “work harder when they know a peer-assessment is coming up”. This is why, for example, when the students had built their activity stations, teacher A “asked other students to test the activity stations too before inviting students from outside of their class to use the activities”. Teacher A continues that “Qridi digital tool is good for peer-assessment”, but that “the students seem to like the non-digital method the best”. Teacher A believes that because “everyone has their own devices and they use them so much anyways that using digital tools for assessment in school does not have any special attraction to it.” Students from school A say they receive feedback “especially from students in [their] own group”. This is the case for both the excellent and poor artifact groups. Students from the excellent artifact group say that when giving peer-feedback as a group to another group, “some groups had some disagreements but reached an agreement”, but they think “giving peer-feedback is nice and easy”, and that “peer-feedback mostly helps them”. In school B, student groups are guided to discuss their work after each lesson, and mark their evaluations in Qridi to assess their own groups’ communication skills and group working skills (elaborated in 1.3.2.2 Case study school B). The observers note that students in school B “did not evaluate the work of each others’ groups” due to time constraints, and as such only provided evaluations about their own groups. The student groups themselves confirmed this in separate interviews. The poor artifact producing group said that “the students inside a group gave some peer-feedback to each other but there was not so much peer-feedback between the groups”, and the intermediate and excellent artifact groups each confirmed this in their own interviews.

Although students in both schools work in groups throughout both learning cycles, and the student groups provide assessments of their own groups' performance, and in some cases provide assessments of the other groups, the students themselves are assessed individually and receive individual grades. There are "no group-specific grades or assessments" for students in school B. The students "receive numerical peer feedback" from other students, but this is feedback, and "not a grade or grading per se". Similarly, in school A, students received individual grades, even though they also "received peer feedback as a group, group self-assessment", and teacher A also used their color chart grading system "to help to improve" the groups' activity stations.

Group discussion is stimulated by the teachers of school A and school B "always" and "sometimes" respectively. In school B, "groups discuss relatively little during the observations", mostly related to practical issues such as the division of labour. Discussion occurs mostly during small group work, during which the students "may even give direct feedback to each other". However, the observers note that there are also lessons in which group discussion is encouraged, such as when teacher B opened a discussion by saying "Let's discuss together how the thermal conductivity study of metals carried out in the previous lesson could have been made more scientifically sound and reliable". In school B, teacher B "instructs students to participate in a joint discussion and reflection", for example, by asking student groups "guiding and suggestive questions", if the group seems to be "stuck". In school A, most of the time the main emphasis is on the students' own ideas, but the "teacher guides and leads the students to notice and find solutions", according to the observers. Indeed, Teacher A thinks "the most important tool is discussion", adding that school A has "tried several digital platforms but no platform can replace face-to-face communication, problem-solving and discussion". Teacher A thinks it is important to have "enough time to listen to what the students are talking about in the groups", and to then "take part in the discussion, to ask for rationalisations and challenge and encourage them". In their group interview, the poor artifact producing student group from school A agrees, adding that "when the teacher discusses with them during the working process, it helps students to improve their work".

#### **3.1.2.5. Facilitate self-assessment**

According to the observers, students are always encouraged in school A and sometimes encouraged in school B to be actively involved in assessment. In school A, students complete a self-assessment on Qridi at the end of almost every lesson during the learning cycles, so the subjects of the assessment actually progress together all the time. The observers note that "In this way, students' self-assessment is also continuous and timely", because "the things to be assessed are remembered more clearly after the lesson than, for example, if the assessments were only at the end of the learning cycle". Similarly, in school B, student groups are guided to discuss their work after each lesson, and mark

their evaluation in Qridi. Still, teacher B believes that success criteria for group work “should be improved and communicated better to the students”, and thus intends to “include students in designing the success criteria”. As the students are “so focused on grades”, the intention is to leverage this to the students to focus on group work skills.

The students are aware of the meaning of evaluation in the learning process most of the time in both schools, according to the observations. As noted previously, teacher B from school B says that the “success criteria for group work should be improved and communicated better to the students”, and “include students in designing the success criteria”. The students of school B think that the self-assessment and evaluation of the learning cycles is clear, but they describe the different tasks being assessed rather than the transversal competencies or the assessment criteria. The poor artifact producing group from learning cycle two say that “It was clear [to the students] what the main topics of assessment are” and “remember that the criteria was discussed”. They say teacher B “gave the assessment criteria [to them]”, which they “think works well”, and that they feel that they “understood the basis of assessment of the project”. The intermediate artifact producing student group say that they feel they “understood the basis of assessment of the project” but according to their example they “know that 30% of grade is based on classroom working, 30% on work report, etc”, but do not discuss the evaluation criteria itself. The students of school A think that they benefit from being reminded regularly of the meaning of evaluation in the learning process. The good artifact producing student group from learning cycle one says “It helps if [teacher A] reminds you along the way about what has been done, what happens now, and what is yet to come”, and although they do “not remember what [they] assessed in Qridi”, they think that “when doing the self-assessment in Qridi, at that moment, helps [them] to redirect [their] actions”. Similarly, the students from the poor artifact producing group “remember that the objectives were presented” and that teacher A “reminds [the students] of a blue objective table that was presented at the beginning of a project, and where it is on the classroom wall”. Further, when asked in their interview, students from the poor artifact producing group remember which skill was targeted in the project, but are not able to specify the skill in more detail. This tracks with the observation that students do a self-assessment in Qridi at the end of almost every lesson during the learning cycles, “so the subjects of the assessment actually go along all the time”. In this way, “students' self-assessment is also continuous and timely”, because “the things to be assessed are easier to remember after the lesson”.

The observers evaluate that the teachers always encourage students to be aware of learning results and help students to introduce changes and improvements in learning in school A, and sometimes in school B. In school A, teacher A says that “students are good at self-assessment, and thinks that “providing the students with assessment criteria helps with this a lot” because “this kind of [self-assessment] skill will not develop just by age but needs to be constantly supported in teaching.” The students from school A largely agree, with both the good and poor artifact producing groups saying in their interviews that working on “projects teaches you skills”, and self-assessing skills in Qridi, or offline, “helps [them] to remember what needs to be done differently or better the next time”, even if “some students are better at this than

others". The observers note that in school B, teacher B encourages students "to find things out for themselves or figure out a task based on their previous experience" in this case in the subject of chemistry. Teacher B refers to the goals of group work and assessment several times, especially in connection with the assignment. In addition, according to the observations, at the beginning of the course the teacher talks to the students about "the importance of learning skills and working life, and learning to learn". Teacher B, from the teacher interview, "would want students to learn to be more independent and skillful in improving their product by assessing and discussing it within the student group", so that the "students would see the value of that process", without teacher B "having to pressure them to do it". However, as the intermediate artifact producing student group from school B explains in their interview, "the work report [from learning cycle one] is submitted at the end of the [learning cycle], and gets graded", but students "do not have the opportunity then to improve it based on the feedback and submit again". Perhaps a more formative, rather than just summative assessment of the work reports, which would give students an opportunity to practise and resubmit, would be beneficial in encouraging students to not just be aware of learning results, but also help students to introduce changes and improvements in their learning.

According to the observers, the students of both schools identify most of the time. what they have learned and what learning they should have accomplished. In school B, the observers note that "students are able to assess learning, especially when looking at the assessment of group work", and "it feels like they have been really critical of the whole group". Similarly, in school A, the observers note that when looking at the groups' self-assessment in the digital tool Qridi, "the groups had rather critically evaluated their own work." Teacher A thinks "the students are good at self-assessment", and can identify how they are learning and progressing. The teacher thinks that "providing the students with assessment criteria helps with this a lot", and that "this kind of [self-assessment] skill will not develop just by age but needs to be constantly supported in teaching." Teacher A says that the students can produce full sentences about what was easy and difficult even though they are only in grade 5. In contrast, the observers note that these answers "could have been more justified and do not always come in full sentences like had been asked".

**Similarly the students attribute learning results to their own efforts and responsibility, and recognise what can improve student learning. For school B, the observers note that there is "great dispersion between the students, but to some extent for everyone, yes."** Even within groups, student self-assessment responses in school B match with the observations. **Indeed, one student group "did not ask for a lot of help"**, but assessed themselves as being in need of help, so "apparently thought that they should have" been asking for help. Similarly, the observers note that in one of the groups, "two students seem to make the effort for the project, and the other two students often spend time chatting and doing other things than what they should", and "this clearly shows in their answers about sticking with the schedule". Apparently these students are capable of reflecting on what they can and should do to improve their own learning, and can attribute the learning results to their own efforts, even when they fall short of their own aspirations.

The observers find it similarly notable that in school B, the student group whose work was chosen as the representative poor artifact example had “assessed their expertise as excellent” in the self-assessment. Even so, the observers indicate that the students of both schools demonstrate reasonable but imperfect competence in assessing their own learning. Observations show that the students of school A “are able to assess learning, but especially when looking at the assessment of group work, it feels like they have been really critical of the whole group.” The good and poor artifact producing student groups from school A both say in their interviews that self-assessing skills in Qridi, or offline, “helps [them] to remember what needs to be done differently or better the next time”, even if “some students are better at this than others”.

The evaluation criteria and rubrics are always provided to the students at the beginning of the learning cycles in both schools according to the observers. At the start of each learning cycle the objectives and assessment criteria are reviewed, to facilitate student evaluation of their own work. However, in school B, the assessment criteria presented to the students may contain only 1-5 scales rather than more in depth descriptions or examples of what actually is required on each of the 5 levels for school B. Indeed, teacher B thinks that “the success criteria for group work should be improved and communicated better to the students”.

The teachers in both schools always encourage their students to autonomously analyse and discuss assessment according to the observers. The teachers always, for school A, or sometimes, for school B, encourage students to be aware of the learning results and help students to improve their learning. In connection with the assessment, the observers note that teacher B and the students go through the self- and group-assessments together, and “evaluate the working methods and the accuracy of the results orally”. Teacher B gives the students feedback in writing, but also orally while the students are doing the research. Teacher B refers to the goals of group work and assessment several times, especially in connection with the assignment. In addition, at the beginning of the learning cycle teacher B “talks to the students about the importance of learning skills and learning to learn”. Teacher B says that they encourage the students, but would “want them to learn to be more independent and skillful in improving their products” by assessing and discussing it within the student groups. In this way teacher B believes that the students would learn to “see the value of that process”, without the teacher “having to pressure them to do it”.

### 3.1.2.6. Help to Elicit evidence of learning

Student participation is encouraged in contributing ideas to discussions and decision making most of the time or always in both of the observation schools, according to the observers. They note that, In school A, “in the activity stations, the main emphasis is on the students' own ideas, the teacher guides and leads the students to notice and find solutions”. When the students from school A were asked in the interviews how do you come up with ideas while working on the project, the good artifact producing group replied through “discussing with classmates” and “searching online”. In school B, contributing ideas to discussion and decision making is encouraged by “asking why the students came to a certain conclusion”, according to the observations.

The learning cycles aim to make every student's contributions necessary to achieve a result, the observations confirm. In the group assessment for school A, students were asked: “Did all members of the group equally participate in the work (division of labor)?” This very clearly prioritises making the result of everyone's work necessary for its achievement. The observers report that “group work seemed very smooth”, and the “division of labor worked well”. Indeed, teacher A's assessment of the artifacts produced in learning cycle one demonstrates the importance of equal contributions to the artifact production. For the good artifact, the “group worked well together and did, for example, discuss the division of labor fluently”; for the intermediate artifact, “there were some challenges at the beginning of the project” related to “discussion of division of labor and equal participation of all”; and for the poor artifact, “the group was not able to share work at the beginning and the construction was left to only two students”, but that “there is one special needs student who does not manage to focus on working together, and this is reflected in the division of labor of the whole group”. The student interviews also highlight that the students from school A know or have learned several ways of distributing work and resolving situations of conflict, including by “voting and lottery”, and that they “know to ask for help from the teacher if [their] own ways do not work”.

In both observation schools, students support each other, explain to each other how to do tasks, and help each other to keep up most of the time. As the observers note in reference to school A, it “depends a lot on the student and the group, but generally the groups work well together and help others”. Teacher A also believes that “some students are able to use the discussed learning goals to guide their actions and even guide their group by reminding them that some things need to be done especially well because those things are being assessed”, even though “there is high variance in this.” The observers similarly note that students have a lot of discussion about what is to be done and how to do it, but here too there are big differences between groups.” Indeed, this is built into the group assessment rubric for school A. Students give their group a grade for working as a group on a four point scale: beginner, moderate, good, and excellent. The rubric for “good” includes “I can often help others” although “Sometimes I need help explaining this in my own words”, and the rubric for excellent includes “I can explain this in my own words” and “If needed, I can teach this to

others.” So, students supporting each other, explaining to each other how to do the task, and helping each other to keep up is inherent to the assessment and self-assessment of group work. Teacher A thinks “it is important that the students give each other peer feedback” as this is another way that they “can learn from each other”. The student interviews from school A confirm this. Students from the good artifact producing group in school A say that peer feedback “mostly helps them”, and students from both the good and poor artifact producing groups say that they “enjoy group work”, and think that “learning in [and from student] groups can happen”. The observers note that students from school B could be seen helping each other by, for example “advising where to find the right research tool”, or “where to look for the right page in the book”. The students’ evaluation of the value of group work in aiding shared student learning in school B was more varied. The poor artifact producing student group believed that teacher B “encourages students to teach other students what they know”, and that they “like working in groups”. However, the intermediate artifact producing group was somewhat more circumspect, believing that “group work skills are important for the future work life” but “working in groups does not necessarily help [them] to learn better in school”. In their interview, the good artifact producing student group discussed a situation in which the group leader had been thinking she was “too dominant” in explaining tasks and dividing the work, but when she asked her group about this, the other students are supportive of her leadership style and said “it is good that she takes the lead”.

The students from both schools take responsibility for their work contributing to the tasks most of the time or sometimes, according to the observations. In school A, the observers note, “some students are really active” and “in a few groups the responsibility is nicely and evenly distributed”, but in some groups “all the work and responsibility falls to a few students”. In school B, it is also observed that in one of the groups, “two students seem to make the effort for the project, and the other two students often spend time chatting and doing other things than what they should”, and “this clearly shows in their [self-assessment] answers about sticking with the schedule”. According to the observers, these students in some sense took responsibility for their failure to contribute to the assigned task in their self-assessment answers. Thus, it is notable that in school B the group assessment evaluation criteria for communication and group working skills included to what extent: “everyone’s thoughts were taken into account in the work”, and “we stayed on schedule”.

The coordination and sharing of task sequences by the students during the learning cycles varies in both schools, depending on the particular task. The observations have assessed both schools as fulfilling this criteria “always, most of the time, and sometimes”, “depending on the task” and depending on how “open or closed” the task is. The observers provide an example from school A in the activity stations, depending on how tasks are divided, “sometimes the teacher guides, sometimes it is left to the students to solve” things for themselves. Similarly the observers assess that in both schools everyone takes responsibility for their work being integrated into the task most of the time. The observers note that in school A, “some students are really active and in a few groups the responsibility is nicely and evenly distributed,

but there are also those groups where all the work and responsibility falls to a few students". As part of their group assessment, students of school A are asked to answer the reflective question, "Did all members of the group equally participate in the work (division of labor)?" Teacher A does "remind students several times during the project about trying to learn versatile things, not only to always want to do something they are already good at in the group", even though the division of labour within the class groups is freely decided by the students themselves. The observers similarly note that within the groups it is possible to divide the work between participants according to each member's strengths. Teacher A's analysis of the artifacts produced in school A note that the level of collaboration in group work ranged from "working well together" and "discussing the division of labor fluently", to having "some challenges at the beginning of the project discussing the division of labor and equal participation of all", but having all students involved in the construction of the activity station, through to being unable to share work at the beginning so that "construction was left to only two students". The observers further report that "group work seemed very smooth", "the division of labor worked well" overall, and there are "no visible challenges in division of labor or joint decision-making". Although the teacher does intervene, for example, if "any student does not participate in the shared work", this is "more about a lack of focus or concentration", than any student being excluded or feeling left out because they don't get their way. The observers note that In school B, "groups discuss relatively little during the observations", mostly related to practical issues such as the division of labour. Discussion occurs mostly during small group work, during which the students "may even give direct feedback to each other". Many of the student groups "work really quietly and have surprisingly little discussion", yet they "all share their work and do their part". However, in a couple of the groups work was more unequally distributed and there was "more burden on the shoulders of a couple of students or just one student". In another student group, one member "was more active" and "was trusted by the others", even though "all actively did their part".Teacher B's analysis of the artifacts produced in school B note that the level of teamwork in group work ranged from "good teamwork" in which "division of work was fluent, even, and respectful of their strengths", down to poor teamwork in which "one student was responsible for making the video" with "not much discussion about working skills" or division of labour. Indeed, the group assessment criteria for communication and group working skills in school B included to what extent: "we planned to do the work evenly among the group members". The poor artifact producing group from school B say that teacher B "got everyone working and encouraged them to be active" in the project, and that "the task was broad enough for all students to take part in it". The intermediate artifact producing group agrees, even though they think "that is the teacher's job". The good artifact producing group adds that teacher B "helps group work if needed, but if the group works well [they] get to work very independently". In the student interviews the intermediate and good artifact producing groups each elaborate that the division of labour in their respective groups was overseen by one member who was elected as their group leader.

Where there are coordination difficulties or disagreements in the way the tasks are solved, conflict is resolved through "collective resolution as a group" "without needing teacher intervention" in school B, the observers note. In school A,

the observers report that “group work seemed very smooth” and there are “no visible challenges in division of labor or joint decision-making”. Teacher A says that “the students are very well trained to settle disagreements”. At this point it is mostly enough to “tell the group that they just have to settle things as a group on their own”. Teacher A “wants the students to “then explain how they solved the problem” once it has been resolved. Both student interviews from school A clarify that the students “know several ways of distributing work and resolving situations”, including by voting, and via lottery, and that they “know to ask for help from the teachers if their own ways [of resolving conflict] do not work”. In school B, the group assessment criteria for communication and group working skills included to what extent: “we negotiated about the solutions”, “we listened to each other”, and “we spoke alternately”. In keeping with this, the intermediate artifact producing student group from school B believes that “the point of group work is to teach [them] to get along with different types of people, and to make shared decisions”.

The students from both schools hardly ever reflect on the acquisition and advancement of their STEM skills, according to the observations, or at least not as being STEM skills. The students of both schools did however analyse and reflect on the achievement of the learning goals most of the time. The observers note that the students from school A “did not discuss STEM skills per se, but the learning goals (competencies and teacher-led subject-specific)” were discussed in detail. The students of both schools also analyse achievement of learning goals, according to the observations, when “filling in Qridi”. In school A teacher A also asks individual students individually to reflect “why did you answer how you answered” in the Qridi digital assessment tool. The poor artifact producing student group from school B “did not think about the targeted competences [group working skills and communication] when they worked.” At the end of their interview, these students are not able to verbalise what skills they have learned during the project when asked to describe with their own words, but when the interviewer reminds them of the targeted skills, then they remember and agree that they “have learned those skills”. The good artifact producing group from school B said that “the questions in Qridi were clear and made you think about the group work skills”. Interestingly, this group, when asked at the end of the interview, are not able to name skills they learned during the project, even though the students clearly discuss the learning process of the skills during the interview. Similarly, when asked, the poor artifact producing student group from school A remember that “teamwork” was the targeted skill in the project, but are not able to specify the skill in more detail. The good artifact producing student group from school A does not remember what they assessed in Qridi, but they think that “when doing the self-assessment in Qridi, at that moment, helps them to redirect their actions”.

The STEM teaching staff do not jointly assess their student’s acquisition of STEM skills because in both schools a single teacher is responsible for teaching and assessing both learning cycles. Although, in the interview, teacher B from school B expresses interest in collaborating with colleagues to “develop assessment together”, and in the future to “implement shared assessment of interdisciplinary learning modules”.

### 3.1.3. Challenges in digital assessment

#### 3.1.3.1. What are the main difficulties?

The basic challenge is in finding the most suitable use of a particular digital assessment tool, so that it creates extra value compared with traditional, offline or analogue formative assessment. Basically, if the outputs are not being used to inform teaching, as formative assessment should, then digital assessment may not be particularly beneficial.

The Finnish case study schools each utilised the same digital assessment tool, Qridi. Overall, students complemented the platform for reminding them about what is to be done, what is being assessed and the assessment criteria, and in some cases redirecting students' focus. However, neither school made particular use of the data the students created in the platform, even though this could potentially be beneficial, for example, in actively tracking student progress lesson by lesson as charts. Indeed, in their personal interviews, both teachers A and B said they had not used any of this tracking data to inform teaching, which seemingly limits the formative digital assessment's intended purpose of supporting student learning.

In school A teacher A says that "face-to-face discussion and direct feedback can be preferable" to digital tools such as Qridi. Depending on the particular context, this may be the case. As can be seen in Figure 2 in the Introduction, school A does not make use of the colour coding that the students are familiar with from offline assessment, so instead beginner, moderate, good, and excellent must be explained with text alone. Confusingly, the black dots in the assessment table in Figure 2 do not represent assessment data points, but are just random points on the screen that each student clicked on as the location to input the justifications for the formative assessment at the end of each lesson. When the digital assessment tool is just used as a more impersonal substitute for existing summative assessment methods it is not adding value. The observers also note that Teacher A thinks "each of online and offline assessment can work", but is unsure if digital is significantly better. The observers think that "based on experience, [teacher A's] students have the same level of output digitally" as for analogue methods. In school A, the class size is quite large, with 30 students. In this context, the observers note that "giving direct oral feedback in class" may be more efficient than doing so "electronically, which takes more time if you want to be genuinely supportive of learning". According to the observers, the students "miss and like" the colour coding assessment method they use offline, and "students do not per se feel that the tool helped". Showing the learning goals in the Qridi platform can be helpful in some contexts, but the goals are also visible at all times on the classroom wall.

In this sense, the use of the digital tool in school B is somewhat more appropriate because it plots the group assessment inputs as a time series, showing progress over time, as shown in Figure 3 in the Introduction. This at least allows the

teacher and students to track changes in their formative assessment over time, which is potentially useful information for the students and teachers. However, the effectiveness of this feedback depends on what students and teachers do with it, even if this clearer presentation of the data gives it a greater chance of being effective. Indeed, teacher B in the interview said that this progress tracking information was not used at all for these learning cycles, but could be beneficial to make use of in the future. According to the observers, Teacher B “thinks online assessment does bring added value to the students”.

The primary concern is that, to make use of the features of a particular digital assessment tool to support formative assessment, it still needs to be used in a considered way by teachers to support the essential reasons for formative assessment, namely to lead to further learning. The data needs to be used to inform teaching, as formative assessment should, if digital assessment is to be used in an advantageous way.

#### **3.1.3.2. Proposals and recommendations from the agents involved**

Both the teacher and students from school B think there could have been more focus on formative assessment and feedback. In their group interview, the student group that produced the intermediate artifact in learning cycle two commented that “the work report [from learning cycle one] is submitted at the end of the [learning cycle], and gets graded”, but students “do not have the opportunity then to improve it based on the feedback and submit again”. A more formative, rather than just summative assessment of the work reports, giving students the opportunity to practise and resubmit, would help students to introduce changes and improvements in their learning. Indeed, teacher B had “wanted to discuss with each group after each lesson to give detailed feedback to help them but did not have time”, but understands that it “would be beneficial” and wants to focus more on formative assessment in the future.

Teacher B from school B thinks that the assessment criteria and rubrics for the transversal competences could be improved to be clearer and more focused. This would make the learning goals and the method for the achievement of the competencies more transparent. Indeed, teacher B thinks that “the success criteria for group work should be improved and communicated better to the students”. Presenting the students with clearer assessment criteria that make the learning goals understandable would also be a helpful method of teaching the students the transversal competencies in question. Teacher B already does something similar in providing clear examples when presenting “good model answers of previous students who have reflected on the mistakes well” in their own final reports, which “also reflect on the mistakes and what was learned because of the mistakes”.

## 4. CONCLUSIONS AND PROPOSALS

The assessment criteria and rubrics for the transversal competences could be improved to be clearer and more focused. This would make the learning goals and the method for the achievement of the competencies more transparent. If teamwork is the transversal competency to be formatively assessed in the learning cycles, then the objectives and assessment criteria should be developed in a considered way to specifically target teamwork skills.

Similarly, the questions that the students answer in the digital assessment tool, in this case Qridi, should be carefully considered. The number of questions to be answered at the end of each lesson should be daunting, and particularly the focus or quality of these questions could be more directly targeted towards formatively assessing the transversal competency. If the questions are not all clearly linked to the targeted competence, then the formative assessment is not valid for its intended purpose of supporting student learning.

Further, there could be greater opportunity for reflection on the students' answers in the digital assessment tool. For example, the self-assessment completed at the end of a lesson could be revisited at the beginning of the next lesson to focus the students' attention on what they can do better, or improve on, in the current lesson. Similarly, there could be more reflection between the learning cycles and at the end of the project, for the students as well as teachers, to focus on the main goals of formative assessment, since learning of the targeted competence continues even after the project is complete. Although there was space during the project for teacher reflection on the use of the digital tool for formative assessment of competences between the two learning cycles, that did not lead to changes in the formative assessment in the Finnish case study schools, and continued in the same way during the second learning cycle. As the background materials of the ATS STEM project clearly highlight, for formative assessment to be valid, it must lead to further learning. Any formative assessment that does not support student learning, is not fulfilling its intended purpose.

In Finland, formal written lesson plans are not standard. However, more detailed planning of the lessons and of the learning cycles could be beneficial in maintaining focus on the learning goals and on the transversal competencies being formatively assessed. This would be especially important for projects and learning cycles being co-planned, co-taught, and co-assessed by several teachers.

An additional challenge in planning, teaching and assessing these learning cycles in schools is the number of components involved. It is not easy or straightforward to plan two learning cycles of lessons when it involves: identifying a topic based on the SDGs; using a STEM framework; selecting and using digital tools that are to target formative assessment; taking into account a local curriculum, in this case the Finnish national curriculum, and its requirements, its subjects, and subject-specific goals; and identifying a suitable transversal competency to target. There are a lot of

moving parts, and each part impacts on the others, so different choices, whether of topics, digital tools, subjects, transversal competencies, learning goals, or the like, inevitably prioritises some parts over the others. This is particularly so given the limited time and resources available to the participating schools and teachers. In the Finnish case study schools, the choice of SDGs, or relating the chosen topic of the learning cycles back to the SDGs, seems to be almost an afterthought, or simply a post-hoc justification, but that is understandable given the large number of components to this project. Potentially a decision tree or some other method of prioritisation could help teachers to decide the most important things first and then work down to the least important components. This would have the added benefit of being standardised across participating countries, making the different implementations more comparable.

The choice of digital assessment in general and the particular method of digital assessment needs to be considered carefully. Digital assessment should be used in a deliberate way by teachers to support the essential reasons for formative assessment, namely to lead to further learning. For digital assessment to be used in an advantageous way, its features and the data produced need to be used to inform teaching, as formative assessment should. Otherwise, transporting effective offline assessment methods to a digital platform can actually be less effective than the original.

One limitation of the study is that in both case study schools the projects were implementations of single teacher projects. As such, and in any case being a study of just two schools from the same municipality, implemented individually by two teachers, they are not necessarily representative of multiple teacher co-planned, co-taught and co-assessed projects. However, in comparison to the other pilot schools in Finland, this is not exceptional. Indeed, in at least five of the eight piloting schools in Finland, the learning cycles were implemented by individual teachers. The autonomy and independence of teachers in the Finnish system can somewhat work against collaboration, and studies, such as OECD (2016), have shown that, particularly in Finland, there is a need to further develop teacher collaboration.<sup>2</sup>

A further limitation relates to the timing of study. The case study schools both implemented their pilot projects at the very end of the semester. This is not necessarily the most suitable time for a demanding project like this, especially given that one of the main goals of the project is to study assessment. By the end of the semester, many grades have already been given, and students are not necessarily so motivated to improve their performance. This can be particularly challenging when formative assessment is being targeted, which essentially is focused on supporting student learning.

---

<sup>2</sup> OECD (2016). Supporting Teacher Professionalism: Insights from TALIS 2013. Paris: OECD Publishing.