Fostering Natural and Data Science Skills of School Kids

Alexander Nussbaumer[1], Christina M. Steiner-Stanitznig[1], Silke Luttenberger[2], Sylvia M. Ebner[1], and Christian Gütl[1]

[1] Graz University of Technology
Graz, Austria
alexander.nussbaumer@tugraz.at
[2] University College of Teacher Education Styria
Graz, Austria

Abstract. While on the labour market there is a high demand for employees educated and skilled in the field of natural sciences, engineering and information technologies, a lack of interest in these fields of expertise can be identified in young people. The aim of our research is to leverage a scientifically founded and empirically validated approach and contribution towards promoting interest and motivation in dealing with data science and scientific and technical subjects. To this end a novel pedagogical approach and a related learning environment has been developed and used in a first pilot phase. This approach was implemented using research and exploration questions on weather forecasts and data from temperature measurements as an example pedagogical scenario. The pilot study was executed in real-world settings of three primary and two secondary schools. The results demonstrate high interest and motivation in the science project and related data science tool.

Keywords: science learning, digital literacy, inquiry-based learning, data science process

1 Introduction

While on the labour market there is a high demand for employees educated and skilled in the field of natural sciences, engineering and information technologies, a lack of interest in these fields of expertise can be identified in young people [5]. Interests of pupils in these subjects and professions is declining steadily from primary level and at the transition to the secondary level. Therefore, there is a need for innovative didactic approaches that are suitable to engage and motivate young people.

This paper describes an initiative that aims at stimulating young people's interest and motivation in dealing with natural science topics and thereby shall foster competence development in scientific and digital literacy. In contrast to other Science, Technology, Engeneering, and Mathematics (STEM) education approaches, such as the use of online laboratories [10][6], our approach focuses on the integration of real world natural science experiments with data science activities.

While learning and teaching data science is well established at university level, there are almost no pedagogical approaches to teach and learn basic data science principles in primary and secondary schools. This paper presents a pedagogical model and an application to teach and learn basic data science skill at schools.

This approach is elaborated and implemented for different thematic areas and empirically deployed and tested in the context of primary and secondary schools. Piloting phases involving trained assistant teachers enable an iterative evaluation and refinement of the approach. The digital learning environment and didactic concept with accompanying material will be provided to schools for application in educational practice.

This paper is structured as follows: The novel pedagogical approach is described in Section 2. An application to a concrete pedagogical scenario with a supporting learning environment is described in Section 3. The evaluation of the first pilot study and respective results are reported in Section 4. In Seciton 5 a conclusion and outlook to future work is provided.

2 Theoretical Foundation

The key approach proposed in this paper to teach and learn data science skills is based on the combination of data science with natural science. While almost no pedagogical approaches are available to learning data science at primary and secondary school level, a lot of pedagogical research exists for learning natural sciences at this age.

2.1 Inquiry-based Learning

A pedagogical approach often applied in the field of science teaching is Inquiry-based Learning. Anderson [1] proposes to apply scientific inquiry in the learning context. Scientific inquiry refers to the diverse ways how scientists study and explain the natural world. Applying this process to the learning context leads to an active learning with distinct learning activities. In general, Anderson reports positive effects of this way of learning and teaching on student achievements.

Four levels of inquiry are distinguished in the literature, according to the degree the teacher is involved in the learning process [2][9]. In level 0 (*verification*) the teacher provides the research questions, the data, and supports the interpretation. In level 1 (*structured*) the teacher provides the research question and data, but the student is responsible for the interpretation. In level 2 (*guided*) the teacher provides the research question and the student is responsible for data collection and interpretation of the results. In level 3 (*open*) the student is responsible for the research question, data collection and result interpretation. In a quantitative study [2] it turned out that especially level 2 approaches (*guided*) of inquiry instruction have positive effects on student learning.

A method to implement inquiry-based learning in science teaching is the use of the 5E Model [3][9]. The 5E model describes five subsequent activities (phases) that students should perform in science learning. In this way it provides a template or pattern for lesson planning, and supports the teacher in the science teaching process. In detail the activities (or phases) are:

- *Engagement.* In this phase the student's interest is stimulated. Questions are raised and connection to prior knowledge is established.
- *Exploration.* In the exploration phase students begin to investigate a problem, pose real-world questions, and develop hypotheses. Furthermore, they perform scientific and laboratory tasks.
- *Explanation.* In this phase the students develop explanations for their observations. They discuss what they have learned and internalise what they have learned.
- *Extensions.* In this phase students generalise their understanding and transfer it to real knowledge, in order to understand real-world problems.
- *Evaluation.* In this phase students compare their previous knowledge with the new knowledge. This phase increases the depth of understanding as students have to apply metacognitive skills.

2.2 Data Science Process

Data Science is an emerging field in computer science and used for many applications. Typically, it is employed to answer real-world problems by using large amount of data. These data are of different type (e.g. structured, unstructured, text). Data Science is applied in several steps, which is described as the data science process [4]. This process consists of the following steps:

- *Research goal.* In this step the problem to be solved is framed.
- *Retrieving data.* In this step the raw data are collected.
- *Data preparation.* In this step the raw data are cleaned and prepared. For example, questions on the time zone, missing data, data range are dealt with.
- *Data exploration.* In this step the data is explored in order to get a first overview and insights. Initial trends can be seen and a first relation to the research goal is made.
- *Data modelling.* In this step a in-depth analysis is performed, such as machine learning or statistical algorithms.
- *Presentation and communication.* In this step the results from the analysis are documented and shared.

2.3 Data Science Learning Model

The combination of Inquiry-based Learning (i.e. the 5E Model) and the Data Science Process, as outlined above, build the foundation for a new pedagogical approch. This approach consists of five steps, whereby all activities or steps from the 5E model and the Data Science Process are mapped and integrated. An overview of this approach is depicted in Figure 1. In detail these steps are:

- ***Engagement and research question.*** In this step the student is presented with the science problem and asked to take up a pre-defined research question.
- ***Data collection.*** In this step data is collected in the real world and entered in a computer system.
- ***Data organisation.*** In this step the collected data is reviewed and organised, which includes filtering the data according to criteria such as time range. Furthermore, students get an overview of the available data.
- ***Data exploration.*** In this step data is interactively and visually explored. Interactive data visualisations and diagrams help to understand the data and to answer the research question.
- ***Research result and presentation.*** The answers to the research questions are presented and conclusions on the overall research goal are drawn.

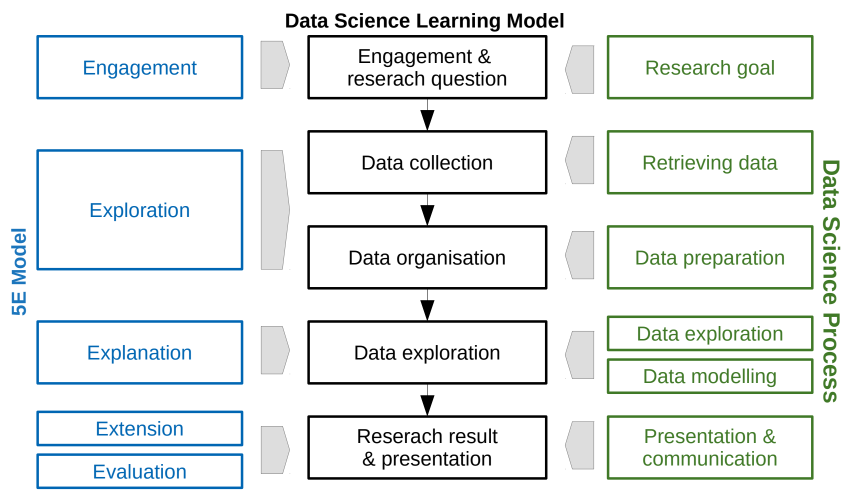

Fig. 1. This figure depicts the Data Science Learning Model with steps and activities integrated from the 5E Model and the Data Science Process.

Since previous research has shown that a guided approach to Inquiry-based instruction is most effective, meaning that a teacher should provide the initial research question, but the student should independently collect and explore data. This concept is also applied to the our Data Science Learning Model application, where we suggest a teacher providing information only for the first step (*research questions*).

3 Application and Learning Environment

This section describes how the Data Science Learning Model for learning basic data science concepts has been applied. This application consists of two parts.

First, it is explained how the science goal and science context is included in a scenario and how this is embedded in the procedure of the pedagogical approach. Second, the learning environment is presented that was used for the pupils to deal with the data.

3.1 Scenario and Procedure

The domain of meterology has been chosen for translating the data science learning model into a concrete pedagogical scenario and application. In detail, the field of weather forecast and testing its accuracy was the theme of the experiment.

In the first step (*engagement and research question*) the children are provided with a weather forecast of the upcoming week, which included the minimum and maximum temperature for each day, as well as the forecasted weather condition (e.g. sunny, cloudy, rainy). A teaching assistant discusses the research question with the children. The research goal is to explore and prove to which extent the forecast is reliable and and whether accuracy decreases the further you get into the future.

In the second step (*data collection*) the children have to collect weather information several times a day. For this reason they are equipped with a micro:bit (see Figure 2) to measure the current temperature. Furthermore, they collect information on the weather condition (sunny, partially cloudy, cloudy, rainy, snowy) by observing the sky. They enter this information in the learning environment together with the information about date, time, and measurement instrument (e.g micro:bit).

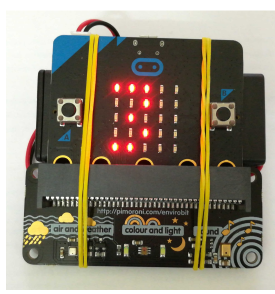

Fig. 2. This figure shows the micro:bit together with the enviro:bit and a battery pack, as it can be used for measuring temperature.

In the third step (*data organisation*) the children get an overview of the collected data in the learning environment. The number of entered data sets (measuring points including temperature and weather condition) is shown on a time range. They can filter, remove, and update measuring points in case they contain unrealistic or false information.

In the forth (*data exploration*) step the children are provided with an interactive visualisation that displays the collected measuring points and the original weather forecast in the same diagram. The children can interactively switch between week view and day view, and they can hide or show minima, maxima, and collected measuring points. This visualisation is used to answer the research question on the forecast accuracy. Details on the visualisation are explained below.

In the fifth step (*research result and presentation*) the children present the result of their conclusion on the forecast accuracy. They are presented with specific questions which they answer in a kind of presentation. In order to create this presentation, they can download specific views of the interactive visualisation.

3.2 Learning Environment

The learning environment is the tool for the children to develop and improve their data science skills. It allows to collect, organise, and explore data, as described in the pedagogical approach (step 2–4). Thus it supports children to establish relations between the natural world, their observations, and research questions on the one hand, and collected data on the other hand.

Besides data collection and data organisation as described in Section 3.1, the key functionality is the data exploration tool. This tool allows to interactively explore the collected data on a time axis for one week (see Figure 3). The measured temperatures are depicted as yellow dots that can be clicked for getting more details (time, weather condition, measurement instrument). On the same visualisation the forecasted minima and maxima are displayed as blue and red bars. This visualisation makes it easy to see if the measured temperature is within the forecast range. On top of the visualisation there are buttons to show or hide the temperature data, as well as the forecasted minima and maxima. By clicking on the weekday on the bottom, the respective weekday view is displayed (see Figure 4). The day view visualisation shows the measured temperature on a time axis for one day and provides the same interactive features as the week view visualisation.

The learning environment provides two further main functionalities. First, it provides a user registration and management tool. Users can register either alone or as group. During the registration process, the children can select an indiviudal icon and a name. Both an own icon and name is an instrument to stimulate motivation through a simple personalisation. Providing support for learning groups is also a method to increase motivation and learning outcome.

The second main functionality is the collection of log data. Each activity a child performs with the learning activity is recorded. Recorded activities include login, logout, enter new data set, select week view, select day view, request data set detail, show or hide forecast. These data are used for activity analysis (see Section 4) and for feeding the teacher tool. The teacher tool (see Figure 5) provides an overview of the activities of all learners or learning groups. It displays how many data sets have been entered and how many activities have been performed in the learning environment. This allows the teacher to trace the activities and quickly detect if learners or learning groups are inactive and need further support.

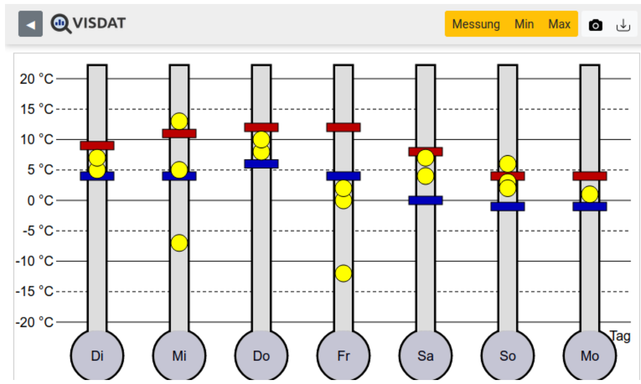

Fig. 3. This image shows the measuring points in combination with the weather forecast for one week.

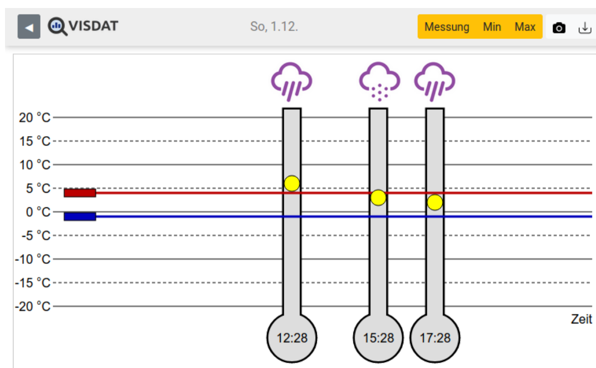

Fig. 4. This image shows the measuring points in combination with the weather forecast for one day.

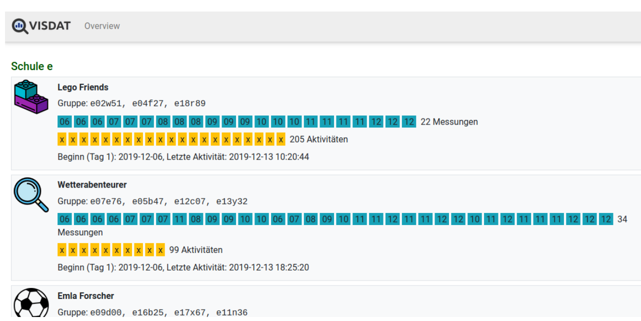

Fig. 5. This image shows the teacher tool with an overview of learning groups and their activities (number of collected data, number of performed activities).

The learning environment is implemented as Web application with a service in the back-end that stores all data. The Web application follows the responsive design paradigm and can be accessed with any Web browser on any device. It is important to note that the learning system does not collect real names or other information that can be used to identify indiviudal children. All collected information is anonymous and adheres to the GDPR.

4 Evaluation

The Data Science Learning Model (Section 2) and its implementation in the example pedagogical scenario and application (Section 3) was deployed and tested in a school context in November and December 2019.

4.1 Participants and Procedure

Children from five different schools (two primary schools and three secondary schools) participated in the evaluation. At each school two classes attended the evaluation, which resulted in ten classes. At both school types third-year grade classes were involved, which means that children at the primary schools were app. 8 years old and children at the secondary schools were around 12 years old.

The total number of children were 173, where 70 children were from primary schools and 103 children from secondary schools. All children were assigned to learning groups of three or four children per group.

The pedagogical intervention and evaluation took place during school lessons. The pilot covered 3 sessions spread over a total of 1 week and were managed by teaching assistants. At day 1 the children were provided with a weather forecast, the research goal, general weather information, and explanations how to collect data in the following week. After a few days the teaching assistants visited the classes again and helped them to enter the collected data in the learning environment in case of problems. In the final session the results were discussed and the learning groups drew conclusions on their research questions in final group presentations. In this way they followed the whole pedagogical approach. Feedback about the learning experience was collected from the school children as well as from the teaching assistants.

4.2 Interest

During the sessions, children had been asked about their situational interest on the different learning activities. The children indicated their current interest on a Likert scale from 1 (low interest) to 5 (high interest). The results are depicted in Figure 6. As can be seen, a high interest could be identified throughout all aspects of the pedagogical scenario in primary/secondary schools (use of microbit: M=4.8/4.1, use of learning environment: M=4.45/4.1, data collection: M=4.1/3.45). Thereby primary school children had higher interest than secondary school children and interest in the micro:bit was higher than in the learning environment and in the data collection process.

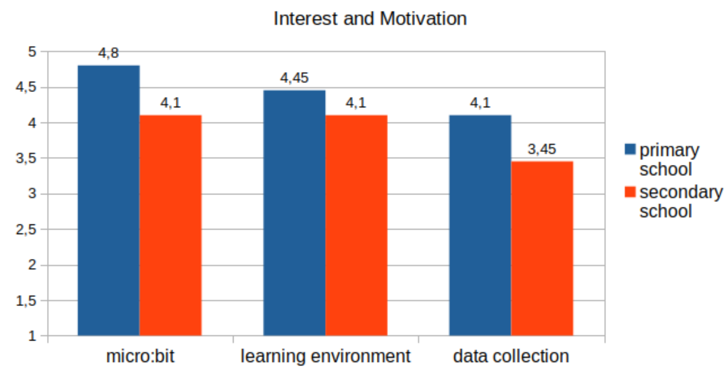

Fig. 6. This diagram shows the average interest and motivation of the children during the evaluation.

4.3 User Experience and Usability

For evaluating user experience with the learning environment in general and with respect to usability aspects, two general questions on enjoyment and an adapted and shortened version of the User Experience Questionnaire (UEQ - based on [8]) was used. In total 6 items (each item consisting of a pair of terms with opposite meanings - e.g. good - bad) of the subscales 'attractiveness', 'perscpicuity' and 'stimulation' from the version for children and adolsecents [7] were used with a 5-point Likert scale ranging from -2 to +2. The results from the UEQ are shown in Figure 7 and represent a quite differentiated picture for the two school types. Primary school children perceived the learning environment very positively on all three aspects, while secondary school children were rather critical in their assessment, especially with respect to attractiveness. This may suggest that the visual appearance and design of the learning environment is more appropriate for younger children. The results on items gathering general assessments on enjoyment (possible score range 1-5) showed that the pupils in general enjoyed the work with the learning environment (M=4.58, SD=1.03 for primary school; M=3.69, SD=1.27 for secondary school) and that they would recommend the overall learning experience to others (M=4.3, SD=1.09 for primary school; M=3.36, SD=1.25 for secondary school), but again with better scores for primary school children.

Feedback from teaching assistants on the experience and use of the learning environment showed quite consistent assessments independent of a primary or secondary school setting. Results are therefore summarized over school types and presented in Figure 8 (possible score range 1-5). Teaching assistants felt that learners enjoyed working with the learning einvironment the learning environment very much and that it was easy to handle for them. Difficulties in dealing with the app and resulting need for assistance were rather rare.

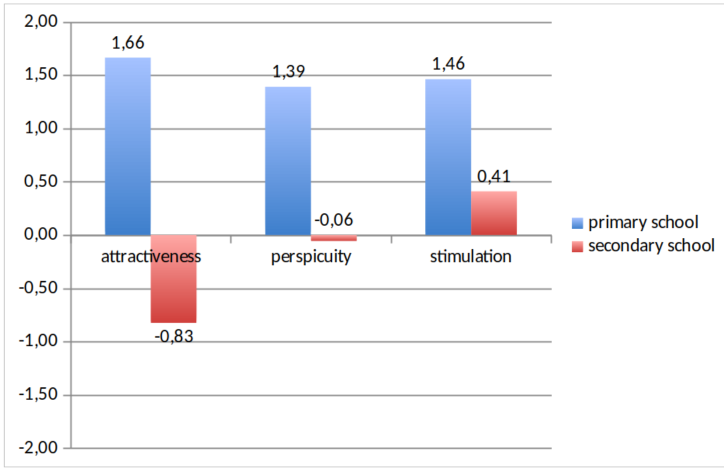

Fig. 7. Results on the subscales of the user experience questionnaire.

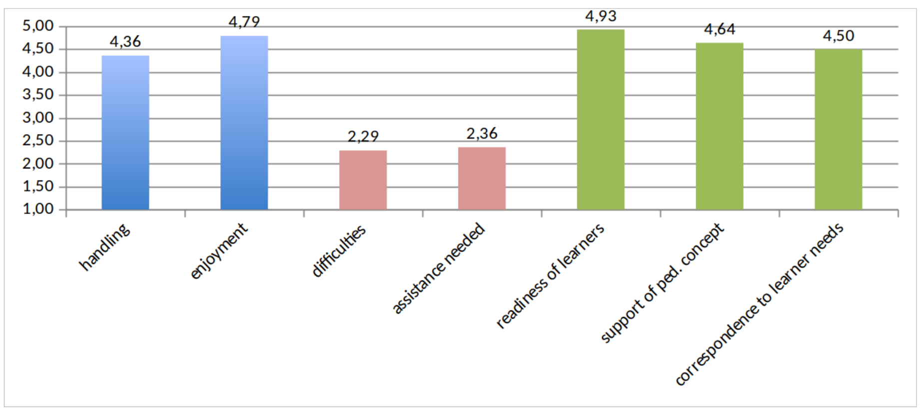

Fig. 8. Overview of results from a teaching assistant perspective.

4.4 Usefulness

From the didactic perspective the learning environment was perceived very well suitable for implementing the data science learning model and did not lack any important functionalities. The feedback obtained from teaching assistants furthermore also highlights childrens' high readiness to work with the learning environment and its good correspondence to learner needs (see Figure 8).

Usefulness for learning had also been assessed from the school childrens' perspective. Children from both school types agreed that they have learned many new things, with primary school children benefitting more from the overall pedagogical intervention (M=4.4, SD=0.93 for primary school, M=3.54, SD=1.08 for secondary schools). Children from both school types indicated that they had understood the learning contents very well (M=4.52, SD=0.68 for primary school, M=4.1, SD=0.96 for secondary school). The evaluation outcomes thus give first evidence that the overall pedagogical approach, in general, and the learning environment, in particular, are a useful and enjoyable approach for learning.

4.5 Activity Analysis

During the pilot sessions the learning environment recorded different types of activities (number of collected weather data sets and interactions with the visualisation module). As learning took place in groups and each group used one computer (mostly tablets), the analysis was made on group level. The result of this analysis suggests high engagement. On average groups at primary schools collected 20.1 data sets (SD=13.1) in one week and groups at secondary schools collected 16.2 data sets (SD=7.7). The average number of all activities is 122.8 (SD=79.1) at primary school level and 110.9 (SD=83.2) at secondary school level. These values demonstrate active participation, whereby children at primary level were slightly more engaged than pupils at secondary level. Furthermore, the high standard deviation value indicates indiviudal differences.

5 Conclusion and Outlook

This paper presented a novel approach to data science learning for school children. While data science is a very complex field, basic steps can be identified and elaborated in a way that school children can be engaged with them. The concept for this approach consists in the integration of Inquiry-based Learning and Data Science Process in the context of science teaching. This integrated approach was used to define individual steps in the learning process. A learning environment supports these steps of the learning process with the help of interactive visualisations. A key feature is the fact that the data to be explored are collected and recorded by learners themselves, which makes them more tangible for children.

An initial deployment and evaluation in educational practice showed high interest and enjoyment of children in the learning activity and the usefulness of the overall approach. User experience of the learning environment showed that the appearance and design seem to be more approriate and therefore more positively perceived by younger learners. This also suggests the considerations of using a different and more sophisticated design and features for older learners.

While the results presented herein focus on the subjective reaction to and perceived usefulness of the learning approach, future work will incorporate also the analysis of effects on data science skills, for a more conclusive demonstration of the effectiveness of the presented approach. Such an assessment will include knowledge and understanding of diagrams, filtering and cleaning data, and understanding different views of data. Currently, the application and evaluation in the context of another pedagogical scenario on a biology topic (plant life cycle) is in peparation.

Acknowledgement

The work reported has been supported by the VISDAT project which has received funding from the Styrian Regional Government in Austria (Land Steiermark), Zukunftsfonds Steiermark, under grant agreement No. 1041.

References

1. Anderson, R.D.: Reforming science teaching: What research says about inquiry. Journal of Science Teacher Education **13**(1), 1–12 (2002). DOI 10.1023/A:1015171124982
2. Blanchard, M., Southerland, S., Osborne, J., Sampson, V., Annetta, L., Granger, E.M.: Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectivenss of guided inquiry and verification laboratory instruction. Science Education **94**(4), 577–616 (2010). DOI 10.1002/sce.20390
3. Bybee, R.: The BSCS 5E instructional model and 21st century skills. BSCS, Colorado Springs, CO (2009). URL https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_073327.pdf
4. Cielen, D., Meysman, A.D.B., Ali, M.: Introducing Data Science. Manning Publications, New York, USA (2016). URL https://www.manning.com/books/introducing-data-science
5. DeWitt, J., Archer, L.: Who aspires to a science career? a comparison of survey responses from primary and secondary school students. International Journal of Science Education **37**(13), 2170–2192 (2015). DOI 10.1080/09500693.2015.1071899
6. Gillet, D., de Jong, T., Sotirou, S., Salzmann, C.: Personalised learning spaces and federated online labs for stem education at school. In: 2013 IEEE Global Engineering Education Conference (EDUCON), pp. 769–773 (2013)
7. Hinderks, A., Schrepp, M., Rauschenberger, M., Olschner, S., Thomaschewski, J.: Konstruktion eines fragebogens für jugendliche personen zur messung der user experience. In: H. Brau, A. Lehmann, K. Petrovic, M. Schroeder (eds.) Usability Professionals, pp. 78–83 (2012)
8. Laugwitz, B., Schrepp, M., Held, T.: Konstruktion eines fragebogens zur messung der user experience von softwareprodukten. In: A. Heinecke, H. Paul (eds.) Mensch & Computer 2006 — Mensch und Computer im Strukturwandel, pp. 125–134. Oldenbourg Verlag (2006)
9. Luttenberger, S., Rath, G., Paechter, M.: Forschendes Lernen. In: U. Fritz, K. Lauermann, M. Paechter, M. Stock, W. Weirer (eds.) Methoden für kompetenzorientierten Unterricht. Budrich, Leverkusen-Opladen (In Press)
10. Maiti, A., Maxwell, A.D., Kist, A.A., Orwin, L.: Merging remote laboratories and enquiry-based learning for STEM education. International Journal of Online and Biomedical Engineering (iJOE) **10**(6), 50–57 (2014). DOI 10.3991/ijoe.v10i6.3997