Education and Information Technologies (2025) 30:18621–18650
https://doi.org/10.1007/s10639-025-13507-7

# Using Jupyter Notebooks as digital assessment tools: An empirical examination of student teachers' attitudes and skills towards digital assessment

Güler Yavuz Temel[1] · Julia Barenthien[1] · Thore Padubrin[1]

Received: 26 April 2024 / Accepted: 20 March 2025 / Published online: 2 April 2025



## Abstract

The integration of different technologies for formative assessment activities into the classroom is very important for the effectiveness of learning and teaching processes. This study is an experimental study in which the student teachers designed jupyter notebooks as formative assessment activities for specified aims and subject contents. For this purpose, student teachers were enabled to design various formative assessment activities with various packages that were developed with the open-access programming languages included in the jupyter notebooks and external applications that were integrated into the jupyter notebooks. We measured the differences in student teachers' self-efficacy beliefs and attitudes towards the design of jupyter notebooks for formative activities before and after the implementation with questionnaires. According to the results of the study, we found positive and statistically significant differences in the self-efficacy beliefs and attitudes of student teachers from various disciplines who participated in the seminar. During the practical activities, we observed student teachers' prejudices and concerns about programming. This process was accompanied by their lack of prior knowledge and prejudices about the open source programming. Despite the difficulties in the learning process, the student teachers designed and developed various jupyter notebooks for different formative assessment activities at the end of the semester.


**Keywords** Formative assessment · Summative assessment · Digital tools · Jupyter notebooks · Attitudes

✉ Güler Yavuz Temel
gueler.yavuz.temel@uni-hamburg.de

Julia Barenthien
julia.barenthien@uni-hamburg.de

Thore Padubrin
thore.padubrin@gmail.com

[1] Faculty of Educational Sciences, University of Hamburg, Von-Melle-Park 8, 20146 Hamburg, Germany

## 1 Introduction

The essential role of assessment, and specifically the formative assessment in improving teaching and learning processes has been emphasized by numerous studies (e.g., Black & William, 1998; OECD, 2005; Praetorius & Charalambous, 2018). The choice and use of an assessment approach depend on the educational purpose. In general, however, standards-based summative assessment, also known as assessment of learning, is often used in classrooms. There is considerable interest in formative assessment strategies, often referred to as "assessment for learning" (Wiliam, 2010a, 2010b). However, the main difference between formative and summative assessment is not in their timing, but rather in their purpose and impact (Harlen & James, 1997). Formative assessment was defined by Black & Wiliam (1998) as follows: "We use the general term assessment to refer to all those activities undertaken by teachers and by their students in assessing themselves that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative when the evidence is actually used to adapt the teaching to meet student needs." As highlighted in this definition, feedback plays a crucial role in enhancing learning and achievement, a point emphasized by various researchers (e.g., Hattie & Timperley, 2007; Price et al., 2010; Vollmeyer & Rheinberg, 2005). For example, Wiliam (2010a, 2010b) identified five key strategies through which assessment can improve the quality of instructional decisions: Clarifying, sharing and understanding learning goals and success criteria, designing classroom activities that elicit evidence of learning, providing feedback to move students forward (feedback). Encouraging students to act as teaching resources for each other, empowering students to take ownership of their own learning. Similarly, Bennett (2011) described these five key strategies as follows: Sharing learning expectations (i.e., clarifying and sharing learning intentions and success criteria), questioning (i.e., engineering effective classroom discussions, questions, and learning tasks that elicit evidence of learning), feedback, self-assessment (i.e., mobilizing students as owners of their own learning), and peer assessment (i.e., mobilizing students as teaching resources for each other). Based on these basic strategies, formative assessment can take different forms. However, although formative assessment can be designed in many different ways to meet different purposes, its main function is to generate feedback on students' performance (Black & Wiliam, 1998; Falchikov, 2005). Through formative assessment, teaching processes can be guided and improved (Wiliam & Thompson, 2008; Ferdig et al., 2020) based on key strategies, where students are, where they are going and how to get them there (through feedback). Several review studies have emphasized that there is a significant relationship between students' achievement and formative assessment strategies (e.g., Black & William, 1998; Hattie & Timperley, 2007; Kingston & Nash, 2011; Palm et al., 2017). However, the effectiveness of formative assessment depends on how it is realized and implemented (Kingston & Nash, 2011; William, 2010a).

Traditionally, educational assessment has been carried out using paper-based methods. In the literature, assessment based on computer applications has been classified into different categories by different researchers. Due to the contributions of technology-based assessment processes, a large number of applications and tools have been developed based on two main classifications (computer-based assessment and computer-adaptive assessment). Many researchers have emphasized the benefits of integrating technology, especially in formative assessment (e.g., Bennett, 2015; Wang & Shin, 2010). The integration of the different technologies to the course assessment offer several benefits for educational providers including paperless test distribution and data collection, greater standardization of test administrations, monitoring of student motivation, obtaining machine-scorable responses for writing and speaking, providing standardized tools for examinees (e.g., calculators and dictionaries), and the opportunity for more interactive question types (Bridgeman, 2009). Some researcher reported also that computer-based tests improve the students' academic performance and motivation (Garrett et al., 2009) and also computer-based tests reduced testing time and developed stronger self-efficacy, intrinsic and social testing motivation in the participants (Chua, 2012).

Because of their contribution, a variety of computer-based tools and applications have been developed to provide feedback to students at various levels and through domain-specific approaches. For example, Moodle (a Web-based learning management system), adaptive content with evidence-based diagnosis (ACED; Shute et al., 2008), supported study (AssiStudy; Rodrigues & Oliveira, 2014) have been developed. Using software or digital assessment systems, students can automatically create tests and receive immediate feedback (Shute & Rahimi, 2017). In addition to these software, many web-based digital assessment tools such as Snappet, Socrative, Kahoot!, Poll Everywhere, Quizizz, Mentimeter and Plickers have been developed.

Many studies have investigated the impact of these digital assessment methods on the educational process (Wash, 2014; Balta et al., 2018; Balta & Tzafilkou, 2019; Fuster-Guilló et al, 2019; Wang & Tahir, 2020; Kohnke & Moorhouse, 2022; Baszuk & Heath, 2020; Zhyhadlo, 2022) and many researchers have conducted comparative studies to assess the effectiveness of these tools (e.g. Gurung et al., 2020; Sullivan et al., 2021; See et al., 2021). Different studies have categorized these tools into a wide range of different applications. For example, a review by Chowdhury et al. (2021) examined different approaches to digital assessment. They classified digital assessment approaches into six categories (online assessments or exams, adaptive assessment, automated assessments, digital badges, recognition of prior learning, and ethical assessments) from the existing literature. Furthermore, several researchers have made suggestions for the effective use of technology in formative assessment practices. For example, in their review study, Børte and Lillejord (2024) identified three criteria for teachers to use technology effectively in formative assessment practices in schools: 1) clear definitions of formative assessment, 2) alignment between digital tools and pedagogical practices, and 3) data literacy to examine and interpret information to improve student learning. Similarly, Bennett (2015) noted that digital assessment requires a coherent strategy that incorporates technology to provide timely feedback and support learning outcomes. Shute and Rahimi (2017) emphasized that providing appropriate and timely feedback to students and personalizing learning are two essential criteria for effective computer-based learning. A wide range of criteria can be categorized to enable teachers to effectively integrate a digital assessment tool into in-class formative assessment activities. These criteria can be summarized as follows: are these tools suitable for quick automated and manipulated feedback, are they easily and freely accessible? In addition, does the integration of these digital tools allow for the implementation of various formative assessment strategies and can they be adapted to different class levels and different content? Another important criterion is that the selected tools can be adapted to or developed specifically for the learning objectives.

Research results also emphasized that the process of selecting the appropriate tool for the learning objectives is important for achievement. No single tool can provide all the conditions needed to develop effective formative assessment approaches (Gurung et al., 2020). In addition, many tools and practices are costly to use, limited in purpose and difficult to upgrade. Using a digital notebook instead of all these tools can provide an opportunity to develop and design assessment strategies for specific learning objectives. This study aimed to design formative assessment activities using jupyter notebooks. Different studies have examined the contribution of jupyter notebooks in the teaching and learning process and their impact on students' attitudes (e.g., motivation, critical thinking, logical reasoning) compared to traditional learning and teaching environments. (e.g., Zastre, 2019; Al-Gahmi et al., 2022; Thomson & Dawes, 2022; Amoudi & Tbaishat, 2023; Lee et al., 2024). Howson (2023) also noted that jupyter notebooks can help reduce the cognitive load for students taking assessments by providing an interactive and visually appealing assessment experience. One of the common highlights of all these studies is that jupyter notebooks can be an effective teaching and learning tool (rich interactive lecture notes, templates for homework and projects, and as a framework for flipped classrooms), but it is also accompanied by difficulties, negative attitudes and prejudices, especially in relation to learning open-access technologies. Furthermore, students may resist or have negative attitudes towards the jupyter application and learning open source software. For example, Hanč et al. (2020) argued that physics student teachers were not prepared for jupyter technology and open analysis, but were positive about using open educational data made available through R Shiny, another open source data science tool. However, Hanč et al. (2020) also stated that physics student teachers can master the basic digital skills to work with this technology, recognizing the huge impact of jupyter technology on their learning, their data and statistics literacy or their professional development. Jupyter notebook (Kluvyer et al., 2016) is a free, open-source platform. Jupyter notebooks offer many advantages for the learning process: They allow, for example, storing and organizing class materials, providing quick access and linking to resources such as audio, video, images, visualizations. One of the most important advantages over other open access tools is that it includes more than hundred programming languages (with a focus on python) (Barba et al., 2019). Therefore, it allows not only the implementation of packages or programs developed in a single programming language, but also the implementation of packages or applications developed in a variety of programming languages or their combinations. Another important advantage is the ease of integrating external systems. For example, Petersohn et al. (2023) used jupyter notebooks as an interface for an external e-assessment system. Similarly, it is also possible to provide feedback to hundreds of students in a very short period of time using nbgrader (Jupyter et al., 2019). Jupyter notebooks also allow the development of one or more formative assessment activities, provide quick and effective feedback, and enable the use of a large number of software packages and external tools. For example, instead of using separate R and python, it allows you to use different programming languages with a single interface. The ability to share and save the results in different digital formats and to analyze the data related to the results are also advantages. Jupyter notebooks make it possible to design and develop a wide range of formative assessment strategies with a single interface, rather than using separate digital tools for different purposes, and with resources available free of charge. Therefore, they have become an important part of teaching and learning processes and assessment approaches, especially in higher education and programming courses. In this study, student teachers designed formative assessment activities (e.g., manual or automatic feedback, quick quiz, projects) for a specific subject content and grade level that they were interested in using jupyter notebooks.

However, especially since the challenges of learning open access technologies, negative attitudes and prejudices accompany this process, student teachers' attitudes, self-efficacy beliefs and prior knowledge are important in the process of designing juypter notebooks for formative assessment activities. Self-efficacy is defined by Bandura and Wessels (1997) as follows:" Self-efficacy is the belief, or confidence, that one can successfully execute a behavior required to produce an outcome such that the higher the level of self-efficacy, the more an individual believes he or she can execute the behavior necessary to obtain a particular outcome". In the literature, it is emphasized that one of the obstacles for student teachers to integrate various technologies into their teaching processes is their various pedagogical beliefs (e.g., Han et al., 2017). Therefore, the importance of implementation experiences and their impact on beliefs and attitudes for the integration of various technologies into classroom activities has been emphasized (e.g., Liu, 2011, 2012; Hsu, 2013; Wang et al., 2014).

Therefore, in this study, student teachers designed jupyter notebooks for different formative assessment activities during one semester. The study evaluated how their prior knowledge, self-efficacy beliefs, and attitudes towards designing jupyter notebooks for formative assessment activities changed with the use of open access resources in formative assessment activities. In the study, three research questions that examined: The first research question was "What is the student teachers' prior knowledge about the functions and applications of jupyter notebooks?". We also examined the differences in student teachers' self-efficacy beliefs about their skills and attitudes after using jupyter notebooks with two research questions: 'How do student teachers' self-efficacy beliefs about their skills in designing jupyter notebooks for formative assessment activities change after the seminar they attended?' and 'How do student teachers' attitudes about using jupyter notebooks as formative assessment activities change after the seminar they attended?'.

### 1.1 Teachers' attitudes regarding the use of digital tools

In addition to cognitive aspects of competences, teachers' competence is assumed to include a range of beliefs, motivation and emotions (Baumert & Kunter, 2013). Beliefs are often defined as assumptions that are subjectively assumed to be true and are consciously or unconsciously held (Voss et al., 2013). They are assumed to play an important role in how students can benefit from teaching in the classroom (Howells, 2018). Such motivational and emotional orientations are of specific importance for teaching with digital tools, because new technological advances are often met with skepticism or reservations. As the implementation of digital tools is one key challenge in education today (Howells, 2018), promoting a positive and realistic view about digital tools and teachers' abilities can therefore be considered a goal of initial teacher education at universities in itself. However, a study involving student teachers has shown that teachers' beliefs regarding digital tools are not constant and develop as they go through phases of purely positive beliefs, disillusionment and differentiated beliefs when learning about digital tools (Speer & Eichler, 2022). Here, targeted courses in initial teacher education or professional development can help to support the development of skills and bring about lasting changes in skills. Indeed, different intervention studies were able to show a positive effect of training courses with digital tools on teachers' beliefs, motivation, and emotions (e.g. Posnanski, 2002; Reinhold et al., 2021; Thurm & Barzel, 2020). For instance, Thurm and Barzel (2020) found a significant impact of the intervention on teacher's technology-related beliefs. Furthermore, the frequency of using technology and the frequency of reflecting technology use increased more in the experimental group compared to the control group (Thurm & Barzel, 2020). Even though few intervention studies focusing on digital tools or technologies in teaching described their interventions and the use of digital tools and technologies in more detail, initial studies that show the effects of interventions on beliefs can provide first indications: For example, in the intervention by Thurm and Barzel (2020) it was pointed out technology supports students' conceptions and teachers were provided with tryout phases. In addition, teachers were encouraged to reflect on the use of technologies. Thereby the reflection focused on the potentials and challenges of using technology for different task in teaching. Core features for effective intervention such as acquisition of content and pedagogical content knowledge, active learning and the duration of the intervention, which have proven to be effective (Garet et al., 2001) in changing teachers' competences, were also identified. Posnanski (2002) also points out the importance of using instruction during interventions that relate to Bandura's sources of efficacy information (1986, 1997) as they can influence the development of self-efficacy beliefs. These sources of efficacy information are (1) performance accomplishments/ mastery experiences (briefly broken down: (repeated) success raises mastery expectancies and builds a stronger self-efficacy), (2) vicarious experiences (self-efficacy beliefs can also be derived from the observation of vicarious experiences; other people can model behavior), (3) verbal persuasion and (4) emotional arousal. In order to bring about a change in self-efficacy beliefs, it is advisable to address these aspects, for example through tryout phases. Overall, despite there being some research on promoting teachers' motivational and emotional orientations regarding digital tools in education, so far, there are few studies and seminars that address student teachers with heterogeneous prerequisites regarding these competences. Therefore, in present study, we focused on teachers' beliefs regarding digital tools and self-efficacy beliefs towards teaching with digital tools in a heterogenous seminar in initial teacher education.

## 2 Research methodology

### 2.1 Instructional procedures

The planning, preparation and implementation process of the study was completed in three stages. First, in the planning phase of the study we identified the learning objectives and developed the teaching–learning process for the seminar. Therefore, we participated in workshops that were offered by the university's data literacy project during the planning phase. During the preparation phase, we wrote python and R code to be used in the seminar and we prepared materials for using the tools to be implemented (e.g., nbgrader). We also developed examples, tutorials and materials for the applications. During the preparation phase, we also prepared the questionnaire utilized in the study. We created presentations and jupyter notebooks to be used during the semester. Three different instructors participated in the study. The first instructor was an expert in R, python programming, jupyter notebooks and assessment. The second instructor specialized in learning and teaching procedures, instructional quality and formative assessment. The third instructor was an expert in python programming. The implementation steps of the seminar were summarized in the flow diagram in Fig. 1.

The steps of the experimental design applied in the study are as follows. As detailed in the participants section, the study was conducted with student teachers who preferred the course in the winter semester at the Faculty of Education. All of the student teachers used the same server system and logged on to jupyter notebooks. Laptops were provided for the student teachers who did not have a computer. Since the semester is 14 weeks long, 14 weeks of planning was done. For each week, notebooks prepared in advance and containing assessment activities and presentations providing the course content were shared with the participants. The issues covered in each week of

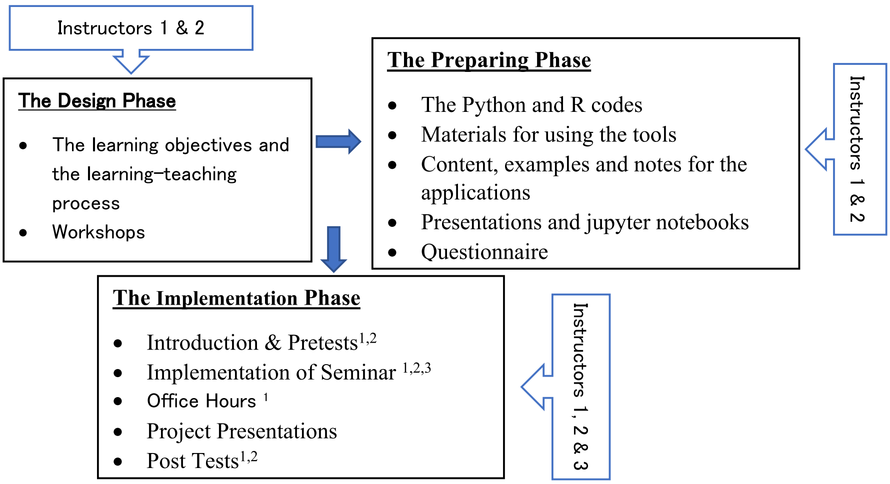


**Fig. 1** A flow diagram of the instructional procedures

the seminar were listed in Table 1. The first three weeks of the fourteen-weeks seminar covered the following topics: introduction to assessment, developing formative and summative assessment approaches, constructing test items. Additionally, we focused the remaining two weeks on basic python programming, two weeks on R programming and one week on creating and designing jupyter notebooks. We used two weeks of the remaining six weeks for nbgrader and feedback and one week for programming open-ended multiple-choice questions. During the last two weeks we practiced peer assessment, mind mapping and e-portfolio, markdown respectively. Furthermore, examples of these designed assessment activities were included in the study (2.4. Designing jupyter notebooks for formative assessment strategies). Since the notebooks of all the student teachers and practitioners participating in the seminar were on the same server, all of the activities were shared with the participants and they were allowed to practice these assessment activities on the notebooks simultaneously. In addition, since student teachers who will be teaching at different levels preferred the course, assessment contents, topics and practices were exemplified for different educational levels and course contents. In addition, we organized question and answer sessions and breakout sessions for the student teachers when they needed them. Thus, every week questions and unclear issues were re-explained. The following sections describe the selection of participants, the research design, the measurement tools used, the design of the notebooks and some sample applications.

## 2.2 Participants

The seminar was declared as an optional seminar for student teachers studying in various disciplines at the faculty of education. The content of the seminar and preliminary information about the planned stages and assessment processes were

**Table 1** The schedule of the seminar

| Week | Thema |
|---|---|
| 1 | Introduction to Seminar, Information about the Procedures, Pre-Tests |
| 2 | Formative and Summative Assessment |
| 3 | Constructing Formative Assessment Activities and Test Items |
| 4 | Introduction to Jupyter Notebooks, Markdown, and Python Programming |
| 5 | Python Programming |
| 6 | R Programming |
| 7 | R Programming |
| 8 | Creating and Working with Jupyter Notebooks |
| 9 | Introduction to the nbgrader |
| 10 | Using nbgrader and Providing Feedback |
| 11 | Quick Tests: Open-Ended, Multiple Choice, True/False Items with Feedback |
| 12 | Peer Assessment and Mind-Mapping |
| 13 | Creating E-Portfolios and Markdown |
| 14 | Presentation of the Projects and Post-Tests |

announced in the content description of the seminar and student teachers from different disciplines preferred to participate in the seminar. The student teachers from heterogeneous disciplines used their preferred subject content and learning objectives for the assessment content and strategies. The seminar was offered as part of a module and the student teachers received formal credit. In the seminar, 32 student teachers participated, however, the sample size for the pre-test post-test comparisons was 19. Comparisons were made on the basis of participants who participated in both pre- and post-tests and who had no missing data. Demographical information of the student teachers who participated in the seminar was presented in Table 2. There were 32 student teachers from various disciplines (sports, history, mathematics, social sciences, German, Islamic religion, biology, English, French, special education /technology, Spanish, music, Latin, chemistry, physics, special education, Protestant religion, geography) participated in the seminar. At the seminar, 19 of the students were males and 13 were females.

### 2.3 Research design and instruments

We used a quasi-experimental (one group pre-posttest) research design. Therefore, student teachers' prior knowledge (e.g., prior knowledge about programming, jupyter), prior skills and attitudes were measured with a pre-test at the beginning of the semester. In the first week of the seminar, before the pretest started, the testing procedure was explained and the test administrator stated that participation in the study was voluntary. An anonymous participation code was created for all student teachers to allow comparison of the pre- and post-tests, considering data protection and privacy in study. In the first week of the study, all student teachers who attended the seminar answered all the questionnaires which are detailed in the next section. Then, in the 14th week of the seminar, the same questionnaires were given to the student teachers again and they filled them in. However, only the responses of the pre-service teachers who participated in both the pre-test and the post-test and who did not have missing data were evaluated in the analyses. Additionally, the assessments activities (which were given in Sect. 2.4.) were practiced every week by all student teachers. However, considering the 14-weeks seminar, some students could not attend some sessions due to illness etc. in some weeks.

#### 2.3.1 Measuring student teachers' self-efficacy beliefs and attitudes

We developed a questionnaire for pre- and post-testing which was applied to all participating students. The questionnaire consisted of four parts. In the first part, we asked questions to measure the demographic information of the student teachers. In the second part, we asked questions to measure their prior knowledge and experiences in programming/digital applications. In the third part, their self-efficacy beliefs regarding the skills to use jupyter technology, programming and digital tools was assessed. In the fourth part, questions to measure student teachers' attitudes towards using jupyter technology and programming and digital tools were administered. The Likert-type scales were developed to measure the construct with a unidimensional structure. The items were measured with

**Table 2** The demographic information of participants

| | Gender | | Target school degree | | | | Semester | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Female | Male | Secondary level I and II | Primary, secondary level I | High level | Special education, secondary I | 1 | 2 | 3 | 5,7 |
| N | 19 | 13 | 4 | 19 | 8 | 1 | 21 | 1 | 8 | 2 |

scale levels between one and five. These scales and questionnaires were provided in the supplementary file. This questionnaire was first used in the pre-test application. The content validity was examined by two researchers who are experts in the field. The pre- post-test Cronbach alpha reliability values of the scale measuring student teachers' skills (19 items) were 0.803 and 0.960, respectively. Similarly, the pretest and posttest reliability values of the attitudes scale (12 items) were 0.723 and 0.607 respectively.

At the end of the semester, the attitudes and self-efficacy beliefs regarding the skills of the student teachers were measured with a post-test by applying the same questionnaire Since the sample size in the study was small and parametric testing assumptions were not met, we decided to use a robust approach for the analysis. Normally, the differences between means can be compared with a paired sample t-test for a quasi-experimental (one group pre-posttest design) research design. We used the paired sample Wilcoxon signed-rank test, which is a nonparametric alternative to the paired sample t-test. However, during the semester, we talked to the students about the challenges they faced during the programming process and provided support during office hours and seminar sessions. To gain an insight into the experiences of the student teachers, in the final presentation of their project the student teachers were asked to address their difficulties with programming. They did a self-reflection on the use of jupyter nootebooks and programming and their applications in classroom.

### 2.4 Designing jupyter notebooks for formative assessment strategies

The different formative assessment approaches described in the literature have different foci and therefore various ways and strategies have been developed to implement formative assessment (Andersson & Palm, 2017). For example, peer and self-assessment approaches can be developed based on the role of students in formative assessment. The main formative assessment strategy is to collect evidence and provide quick feedback to students in order to accurately determine where they are, where they are going, and how they are going in terms of their learning process. One strategy for identifying and correcting learning gaps is to ask short tests or open-ended questions and provide quick feedback on the answers, as well as to report and share the results with the students. Jupyter notebooks were designed for quizzes and different applications and packages were used in this study for feedback. For example, the student teachers developed multiple-choice questions, open-ended questions or front-back-face cards for a preferred subject using python and R simultaneously. The digital tests developed by the student teachers were designed not only with questions and answer options, but also with automatic feedback based on the answer. Moreover, since the jupyter notebooks can be saved in different formats, the digital tests could be quickly saved in various formats (e.g. HTML, pdf) and shared with the students. In a similar way, we used nbgrader, which is also quickly and effectively used for both automatic and manual feedback. For example, nbgrader can be used for peer- or self-assessment as well as for automatic or manual feedback on an assignment or term project. In the following, we describe the various methods for assessment that we addressed in the seminar. As one of the aims of the seminar was to show students the various possibilities and advantages of using digital tools for formative assessment, in particular jupyter, we used assessment methods for two assessment methods that have been shown in empirical studies to be effective for promoting students' learning: peer and teacher assessment (Doubleet al., 2020). To give opportunities for active learning we practiced with the student teachers' the programming of quick tests, homework, mind maps, peer assessment, e-portfolios and feedback. These assessment methods can be supported by the jupyter tool, in which for example, the evaluation of the answers is automated and feedback on these answers is also given automatically. As a result, the student teachers have work to do with the programming, but save themselves the time-consuming correction of tests. To stimulate reflection, the advantages and disadvantages of the individual methods and their implementation with jupyter notebooks were discussed with the student teachers. Additionally, results and feedback can be recorded and quickly shared with students. In addition, data on the results can be summarized in tables and graphs with the packages for the data literacy which included in the jupyter notebooks. We provided some examples of jupyter notebooks which were designed for different formative assessment strategies in the following sections.

#### 2.4.1 Using nbgrader for homeworks and feedbacks:

Nbgrader is a flexible coursework creation and grading tool that can be used for both computer science and for other sciences (Kluvyer et al., 2016; Project Jupyter et al., 2019). Nbgrader has been designed as a system for assigning and grading notebooks since it was designed by Jessica Hamrick in 2014. It has become very popular in education (as of May 2018, there are more than 10,000 nbgrader-based notebooks on GitHub (Project Jupyter et al., 2019). In addition, nbgrader can also be used for homework or to provide feedback after taking a multiple-choice test (e.g., Yuster, 2022). A detailed documentation for the installation and configuration of the nbgrader has been provided by Project Jupyter et al. (2019) (for more information on using nbgrader see also https://github.com/jupyter/nbgrader and https://nbgrader.readthedocs.io/en/stable/). However, a brief introduction to the installation and use of nbgrader followed by an example that was implemented within the scope of the project has been presented respectively.

After installing nbgrader (for installation: pip install nbgrader), a separate tab is available in the jupyter interface. There is also an option to create assignments in the notebook. Nbgrader requires an exchange directory with access rights to the classroom in order to upload and collect the task. In other words, the codes written in the root directory of the nbgrader configuration must be executed in the notebook. The Fig. 2 shows a screenshot of the notebook interface after the installation of nbgrader. Three sub-tabs are available under nbgrader as shown in the figure: Assignment list shows the student interface for sharing, downloading and submitting assignments. Formgrader shows the teacher main interface for assignment management and the course list shows the teachers overview of the course. Create Assignment allows to create an assignment for each cell.

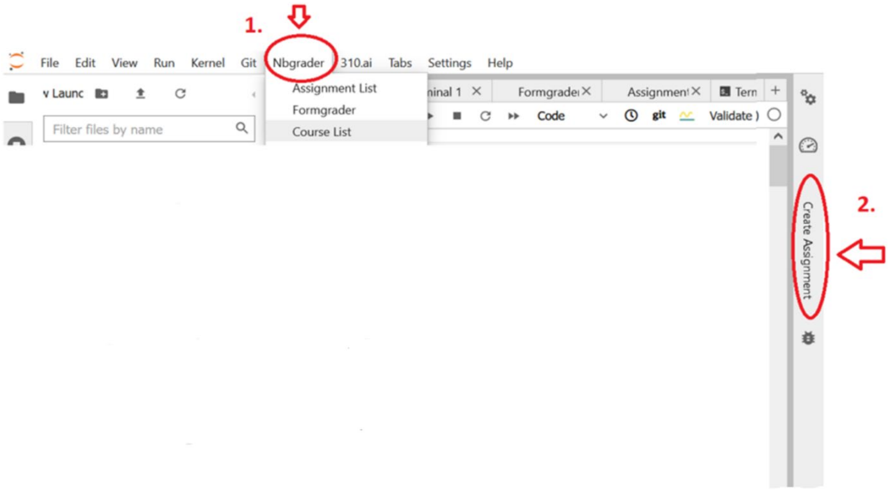


**Fig. 2** Screenshot of the jupyter notebook interface after successfully installing nbgrader

After the installation and configuration of nbgrader, the following steps must be performed in the following order: Create assignment notebook (teachers' version), generate student's version of assignment notebook, release assignment to exchange directory, collect submissions, grade manually or automatically, give feedback. For example, we programmed a simple math question with three simple operations and then an assignment was created for each cell, as in the notebook shown in the screenshot in Fig. 3 below. The next steps were to generate the student version and then previewed how the student version looks with the preview option. There are two ways to submit the prepared assignment notebooks to the students, the first one

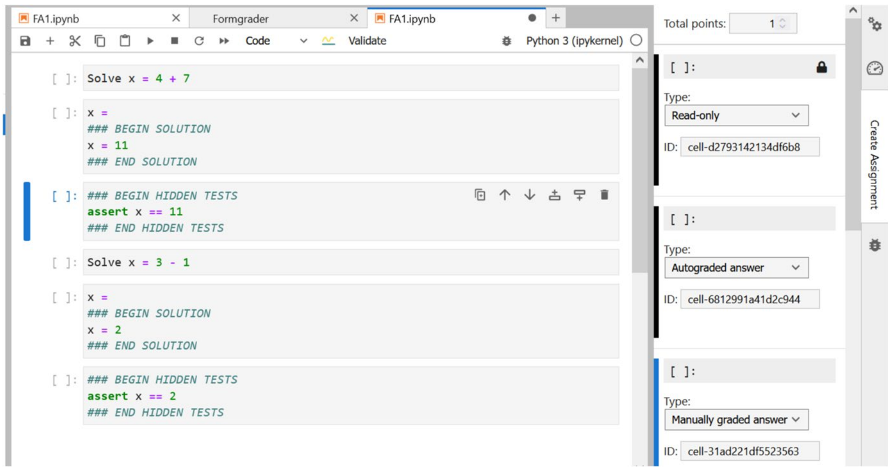


**Fig. 3** Screenshot of the jupyter notebook interface after creating assignments

is by using nbgrader with the "release" and "collect" options, the second one is to do the "release" and "collect" manually (for example, sending it to the students via e-mail and then collecting it from the students via e-mail again).

After collecting the notebooks answered by the students, either manually or through "nbgrader" (with the "collect" option), the student teachers are provided with manual or automatic grades and feedback, and finally the pdf files created by nbgrader for feedback could be shared with the students. Figure 4 below shows a screenshot of an automatic and manual grading and feedback process implemented in the project.

nbgrader has the advantage of being able to give manual or automatic feedback to all students in the class and share the student version of their notebooks in minutes. However, the installation and implementation steps of nbgrader are quite complex, so we have spent a little more time on the installation of nbgrader than on the installation of other application.

### 2.4.2 Create quick tests and provide feedback using the jupyterquiz package

An essential part of assessment and evaluation approaches is to give feedback to students at the end of the course to determine how far they have achieved the learning objective and thus improve learning gaps. For this purpose, two applications were made within the scope of the project. First of all, basic statements (e.g., if, if else, while, for, break, def, global, etc.) necessary for writing functions in python were practiced. The teacher students learned how to write a function at a basic level and applied it, for example, to write a multiple-choice test. The second application was done using the juypterquiz package. Juypterquiz is a tool for displaying interactive self-assessment quizzes in jupyter notebooks and was designed by Shea (2021) (for more information see https://github.com/jmshea/jupyterquiz and https://jmshea.github.io/Foundations-of-Data-Science-with-Python/03-first-data/4-summary-stats.

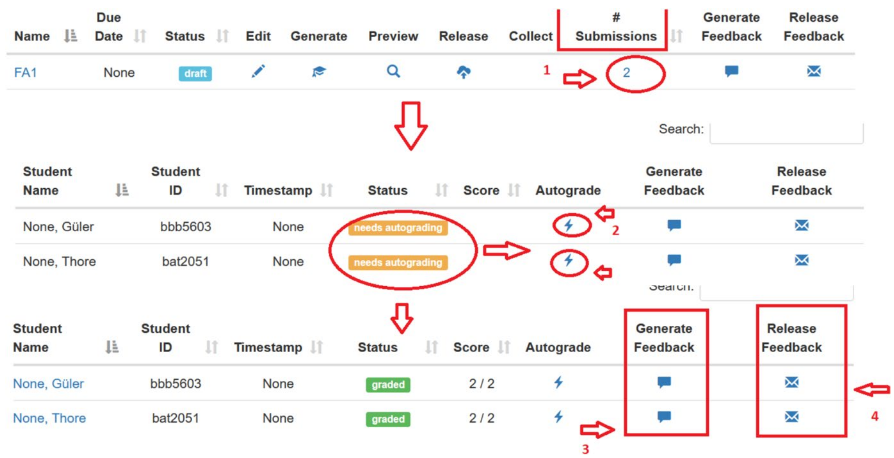


**Fig. 4** Screenshot of the nbgrader interface after submissions and grading

html). Designed as an interactive quiz generator, the package supports two types of quiz questions (multiple choice questions and numerical answer via a text book). It is also possible to manually add different types of feedback to each type of question to determine whether the answer is correct or incorrect and, if incorrect, why it is incorrect.

After preparing the questions, they can be uploaded in different ways (python list of dicts, JSON local file and JSON file via URL). In the project, students from different disciplines prepared questions on a topic that they preferred. Then they wrote the necessary text to give feedback and created a JSON local file through RStudio using jupyter notebook. In the last step, they created the questions and feedback through jupyterquiz. Additionally, they saved their notebooks as "HTML" files with the "save and export notebook as" option provided with jupyterquiz and created an online test application. A screenshot of the JSON local file and the final test form are shown in Figs. 5 and 6.

One of the most important motivating factors in this study was to make jupyter notebooks part of the teaching, learning and assessment process. Designing a test consisting of multiple choice or open-ended items in a digital environment and sharing it with students through a link attracted a lot of interest. However, there were difficulties in programming the open-ended questions. Student teachers considered that this package was more suitable for multiple-choice questions. Some of them programmed the open-ended questions themselves using python instead of developing them with the help of the package. The student teachers practiced the steps of using the package in a very short time. Especially writing packages like jupyterquiz and supporting the process of visualizing the questions can be advantageous for students in the lower grades and for students with test anxiety. The idea of uploading

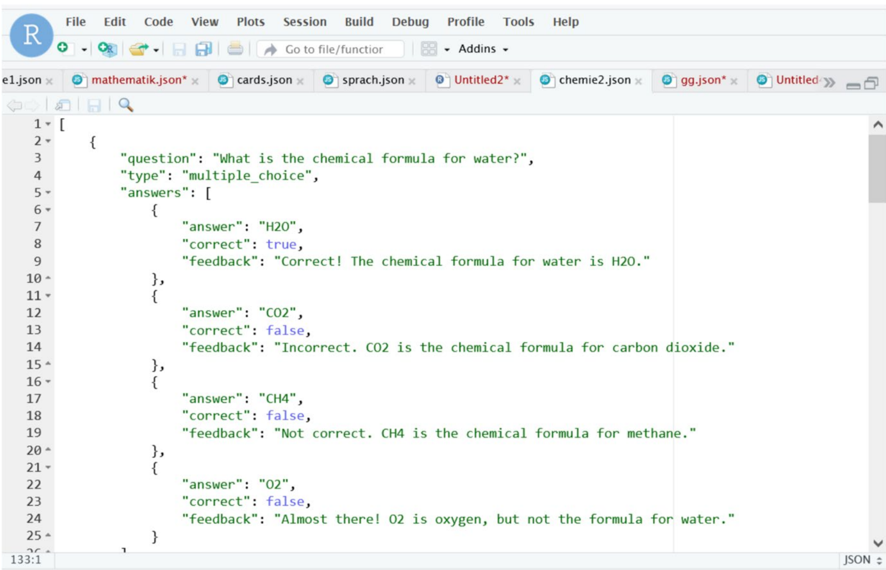


**Fig. 5** A screenshot of the JSON local file

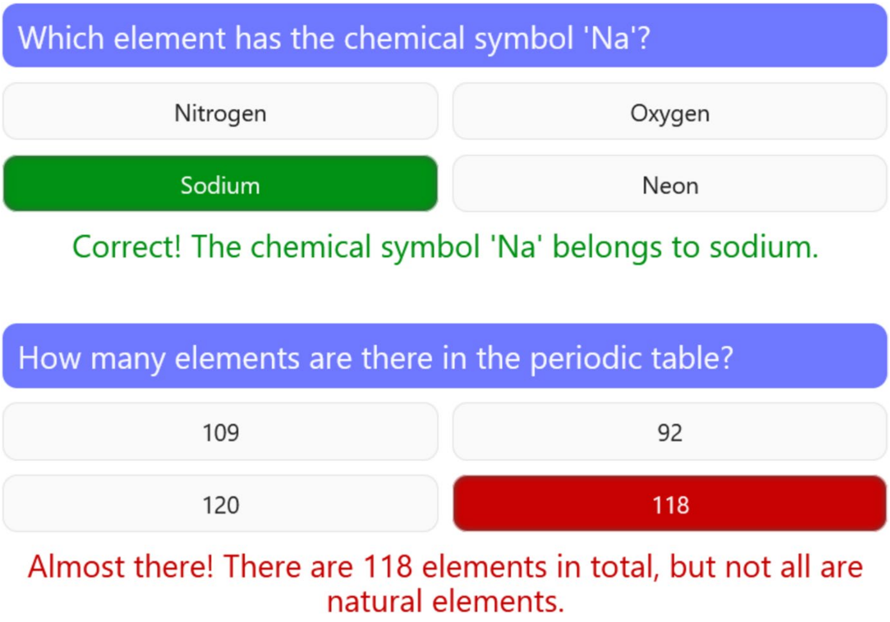


**Fig. 6** The final test form which created with using jupyterquiz

the questions directly to the jupyter notebooks with an existing package instead of programming them directly is very practical.

#### 2.4.3 Create mind maps and provide feedback using the jupytercards package

We used the "Graphviz" package to visualize the network of relationships between complex concepts and to create mind maps. Graphviz is designed for drawing graphs and the package provides two main classes, "graphviz.Graph" and "graphviz.Diagraph". These classes construct graph definitions in DOT language for undirected and directed graphs respectively (for more information see https://graphviz.readthedocs.io/en/stable/manual.html). After installing pydot and graphviz in the python environment (pip install pydot, pip install graphviz), we created the graph, the pydot graph object and the nodes to represent the concepts. In the last step, the created nodes were added to the graph. Screenshot of the codes and mind maps created for an history example provided in Fig. 7. Creating mind maps in this way was particularly advantageous in terms of time. After programming, the student teachers quickly adapted the codes for a new topic. It is very advantageous to create mind maps in this way, especially due to the efficient use of time and the flexibility it provides when creating networks between concepts.

Similarly, we simulated cards that could be used at the end of the lesson to provide small visual reminders or for a quick question-and-answer session. For this purpose, we used the jupyercards python package, created by Shea (2021) in a very similar way to the jupyterquiz package, for the creation of interactive quizzes with

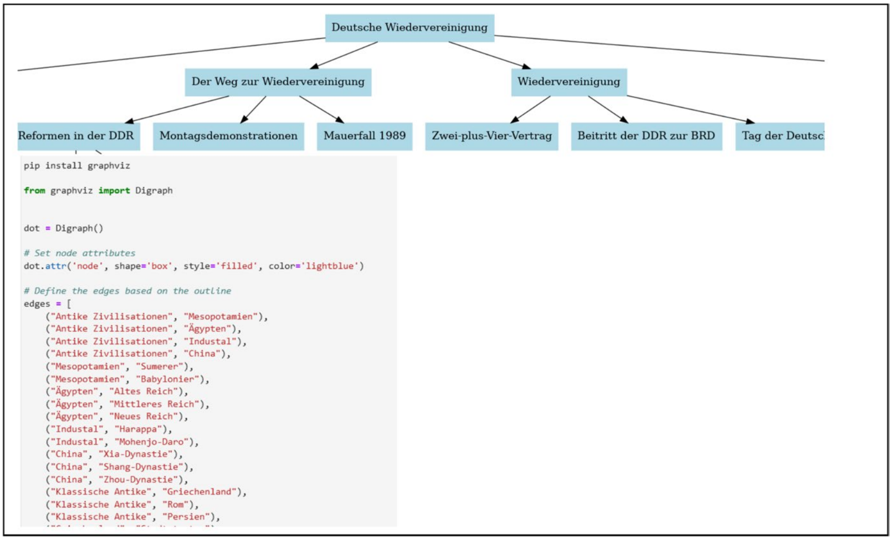


**Fig. 7** A Screenshot of the codes and mind maps created for History

interactive flashcards (for more information see https://github.com/jmshea/jupytercards). Within the seminar, the student teachers designed jupytercards in different subject areas. A sample animated GIF showing jupytercards provided in Fig. 8. The cards are designed so that each animated card has a question on the front and an answer on the back. In addition, with jupytercards, for example, if it is desired to create a card animation with 10 questions and answers at the end of a lesson, it is possible to switch between questions with the next option. Similar to the jupyterquiz package, the flash card can be created in different ways. In the seminar, RStudio was

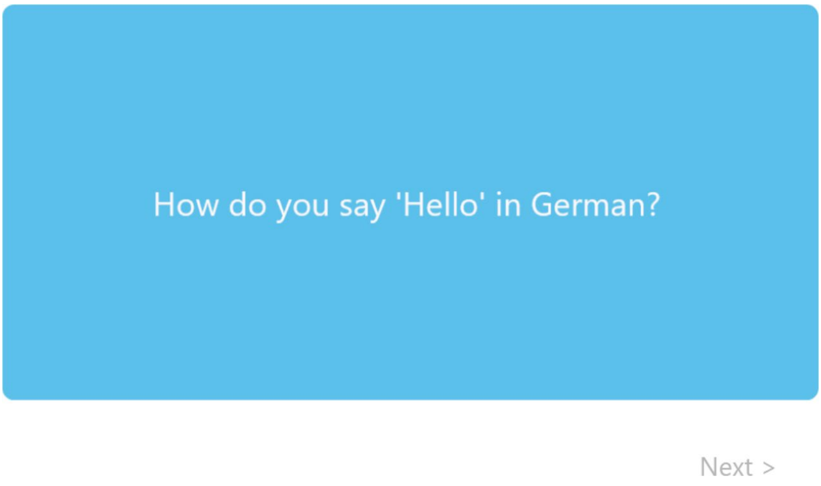


**Fig. 8** A sample animated GIF showing JupyterCards

used to create the questions and answers as JSON data and then the questions were written in the jupytercards package to create the cards.

In particular, one of the most important advantages of open access software is the rich visual content it provides. As the student teachers' adaptation to the programming process has improved, they have started to discover the visual advantages provided by python and R packages in particular. For example, student teachers of history, who had preconceptions about mathematics and computing, created mind maps describing complex historical events. However, the major difficulties experienced in the first weeks of the seminar were gradually reduced by the end of the semester. Regular question and answer sessions were held with the student teachers and instructor 1 and instructor 3 helped the student teachers to solve the problems they had during the programming phase.

The student teachers had no difficulty using the jupytercards package as it works very similarly to the jupyterquiz. Especially the students in language teaching programs were interested in it. They simulated writing words on the front of the cards and explanations on the back and developed it as a learning tool as well as an effective assessment tool.

### 2.4.4 Peer assessment with python

Peer assessment is one of the methods whose importance in formative assessment approaches is often emphasized in the literature. A suitable interface for this approach, which allows students to give feedback to each other in in-class assessment and evaluation applications, has been programmed with python. This application, designed as a simple example, was developed by the students and presented as a project at the end of the semester. A screenshot of a sample application is provided in Fig. 9.

nbgrader was partly used for peer assessment. Moreover, the pre-service teachers also created an example of a notebook for peer assessment using python. They saved

```
[1]: import ipywidgets as widgets

class PeerAssessment:
    def __init__(self, submissions):
        self.submissions = submissions
        self.peer_reviews = {}

    def run_peer_assessment(self):
        for student in self.submissions:
            self.peer_reviews[student] = self.collect_peer_review(student)

    def collect_peer_review(self, student):
        print(f"Peer review for {student}:")
        # Display relevant content for assessment (code cells, text cells, etc.)
        assessment = widgets.Textarea(description=f"Peer Review for {student}")
        display(assessment)
        return assessment.value

# Example Usage:
submissions = ["Student_A", "Student_B", "Student_C"]
peer_assessment = PeerAssessment(submissions)
peer_assessment.run_peer_assessment()
```

**Fig. 9** The screenshot of the code that was written for peer assessment their notebooks as "html" files and shared them with each other. For example, they used criteria such as content, presentation structure, presentation skills, reactions to questions, self-reflection, content organization and creativity, time management, overall impact to evaluate the presentation made by the group members. The feedback that the pre-service teachers gave to each other enabled them to be active in the evaluation process. Although it increased interest and active participation of the students, we also faced challenges, especially because the implementation was more time-consuming and there was no practical tool that could be directly integrated into the notebooks.

### 2.4.5 Create E-potfolio and learning and teaching content with assessment practices using Markdowns

Jupyter notebook markdown can be used as a powerful and flexible tool for preparing learning and teaching materials. Markdown can also provide the opportunity to create well-structured and at the same time visually appealing documents. Different images, external sources, codes, animations or mathematical expressions can be integrated into texts. Markdown also supports headings, lists, tables and other formatting options, making it suitable for organizing content in a clear and consistent way. Since it is possible to save all content as pdf or html, the markdown can be used to prepare an excellent course material that can be shared with students and, if necessary, printed and shared with them on paper. Barba et al. (2019) also stated the advantages that jupyter notebooks can provide to teachers in the teaching–learning process as follows: "Notebooks can be adopted at a variety of levels and formats, offering flexibility based on the needs of a course and comfort/interest level of the instructor: in-class demos, interactive labs, auxiliary material (e.g., book replacements, lecture note supplements), assignments, or full course content in a flipped learning environment. Notebooks offer a route to active learning methods for instructors without experience of them, but do not force a particular teaching style." During the semester, we used markdown for different applications. We prepared a course content using markdown and a screenshot is shown in Fig. 10.

## 3 Results

The study consisted of designing formative assessment activities in the classroom using various open-access packages and applications that are available in jupyter notebooks. The main focus was to measure how student teachers' self-efficacy beliefs and attitudes towards these practices changed over the semester. The questionnaires developed for the project were given in supplementary file and, as can be seen, the first part of the questionnaire assessed students' basic prior knowledge, while the second and third parts assessed students' attitudes and skills regarding digital applications and their contribution to classroom teaching and learning activities. The results obtained from the answers of the students to the questionnaire were presented below in the following order.

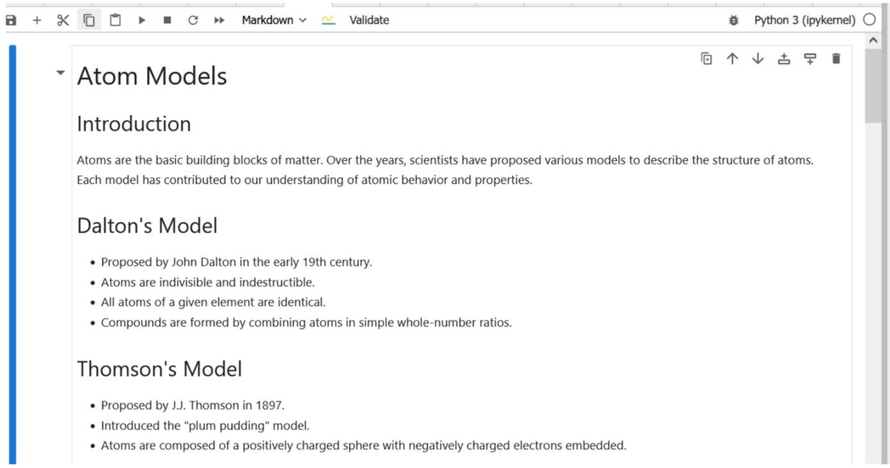


**Fig. 10** Screenshot of a course material prepared by using Markdown

### 3.1 Student teachers' prior knowledge and experience in programming and digital applications

The student teachers' responses to the pre- and post-test questions that assessed their prior knowledge about digital applications with self-reports are provided in Table 3 as below.

Table 3 provided the results regarding student teachers' knowledge and experience in programming/digital applications. The pre-test results showed that the student teachers on average had low prior knowledge, especially about jupyter and other open sources (e.g., python, R, GitHub), but they had experienced or known certain things about digital applications. For example, all of the student teachers had filled out an online test or questionnaire before and almost all of them had heard of digital assessment tools (93.8%) and applied (87.5%) other digital assessment tools before the seminar. This result is consistent with the conversations with the student teachers after the pre-test application as there it was stated that the student teachers had knowledge about general digital applications. They also reported that they had heard of jupyter process and R, python but were unable to use them. On the other hand, the mathematics student teachers were a little bit familiar with programming, but they were mostly able to use programs such as Matlab. In Table 3, this was observed among the answers given to the question "I know what a jupyter is" about jupyter. Approximately 95% of the student teachers who participated in the pre-test stated that they did not know what a jupyter was. In addition, the student teachers who had heard of jupyter stated that they had not used it before. However, although there were differences in terms of their prior knowledge in terms of their programming aptitude, their prior skills in the use of R, Python and jupyter were quite similar: Most of the student teachers had not used them before.

**Table 3** Knowledge and experience in programming/digital applications

| Items | Pre-test | | Post-test | |
|---|---|---|---|---|
| | **Yes (%)** | **No (%)** | **Yes (%)** | **No (%)** |
| I have a notebook with Internet access | 93.80 | 6.30 | 93.80 | 6.30 |
| I have a PC with Internet access | 50.00 | 50.00 | 43.80 | 56.30 |
| I have already attended a course on digital applications in the classroom | 37.50 | 62.50 | 68.80 | 31.30 |
| I know what Jupyter is | 6.30 | 93.80 | 100 | - |
| I know what a digital notebook is | 50.00 | 50.00 | 93.80 | 6.30 |
| I know what R is | 12.50 | 87.50 | 87.50 | 12.50 |
| I know what Python is | 50.00 | 50.00 | 100 | - |
| I have already experienced examples of digital applications in the classroom | 87.50 | 12.50 | 93.80 | 6.30 |
| In my seminars at university, some digital applications have been used in teaching | 87.50 | 12.50 | 87.50 | 12.50 |
| I have heard of some digital assessment tools that have been used in teaching (e.g. Anton App) | 93.80 | 6.30 | 100 | - |
| I have heard of artificial intelligence | 100 | - | 100 | - |
| I know Chat GPT | 100 | - | 100 | - |
| I have used Chat GPT before | 75.00 | 25.00 | 93.80 | 6.30 |
| I have already done an online test or online questionnaire | 100 | - | 100 | - |
| I know what open source software is | 50.00 | 50.00 | 75.00 | 25.00 |
| I know what GitHub is | 25.00 | 75.00 | 56.30 | 43.80 |

**Table 4** Comparison of pre-post test results of attitudes towards the use of digital applications in the classroom with Wilcoxon Signed Rank Test

| Related-Samples Wilcoxon Signed Rank Test Summary | |
|---|---|
| Test Statistic | 17.500 |
| Standard Error | 14.261 |
| Standardized Test Statistic | −1.963 |
| Asymptotic Sig.(2-sided test) | .050* |

**The significance level is .050. Asymptotic significance is displayed*

**Table 5** Comparison of pre-post test results of skills related to the use of digital applications in the classroom with Wilcoxon Signed Rank Test

| Related-Samples Wilcoxon Signed Rank Test Summary | |
|---|---|
| Test Statistic | 125.000 |
| Standard Error | 19.323 |
| Standardized Test Statistic | 2.950 |
| Asymptotic Sig.(2-sided test) | .003* |

**The significance level is .050. Asymptotic significance is displayed*

### 3.2 Student teachers' attitudes and self-efficacy beliefs regarding their skills towards using jupyter notebooks in the classroom

Student teachers' attitudes towards the use of digital applications in the classroom were measured in the second part of the questionnaire in supplementary file and their skills and attitudes towards programming and the use of jupyter notebooks were measured in the third part of the same questionnaire. It was evaluated whether the attitudes and abilities of the student teachers differed after the seminar with the pre-test and post-test applications and if this change was statistically significant. Since the sample size was small, the Wilcoxon signed rank test, a non-parametric alternative to the paired-sample t-test, was used. Tables of pre and post test results were presented below. When the values given in Table 4 were examined, it was seen that the difference in the attitudes of the student teachers was statistically significant ($p = .05$). At the end of the semester, it was measured that the student teachers had a positive attitude ($\mu = 48.000$) and it was observed that the average positive attitude increased compared to the pre-test results ($\mu = 46.438$) and this increase was statistically significant.

Although the statistical result regarding the attitudes of the student teachers was significant, the difference between their means was not very high.

Similarly, when the results given in Table 5 were examined, it was determined that the change in the abilities of the student teachers was statistically significant. According to the results presented here, it is seen that the post-test mean results were higher ($\mu = 58.563$) than the pre-test results ($\mu = 44.630$) and the statistical difference between the pre-test and post-test was significant at the 0.01 level ($p = .003$). One of the most important outcomes was the significant change in student teachers' self-efficacy beliefs regarding their skills. It was observed that even though the student teachers had some prejudices about the use of the jupyter, they improved their skills. This was also clearly observed in the seminar practices. In the first sessions the student teachers needed considerable additional help. Instructor 1 provided examples and applications while instructor 3 moved around the room and helped the student teachers to solve the problems they were having. The difficulties experienced by the student teachers and the number of questions they asked gradually decreased towards the last sessions. The student teachers also overcame their difficulties by asking questions during the scheduled office hours. At the end of the semester, although some student teachers were still hesitant to use them, most of them stated that they could use jupyter notebooks when they needed them.

## 4 Discussions

Digital tools can be helpful for teachers' instructional practice and promote students' learning (Howells, 2018). However, there is often still skepticism towards these tools, teachers feel less self-efficacious and do not recognize the value of the tools for their teaching. For this reason, a seminar for student teachers was developed as part of this study that embeds the use of digital tools in the context of formative assessment relevant to teaching. The practical outcome of the study was that student teachers could develop course-assessment activities as jupyter notebooks. One of the most important advantages of the jupyter notebooks is the opportunity to use pre-existing applications (e.g. python packages or different software tools), but also to allow student teachers to create and develop their own tools or approaches with programming if there is no suitable tool for the purpose. Jupyter notebooks offer a wealth of practical applications thanks to the many programming languages they contain. In this study, it was emphasized that various formative/summative assessment approaches can be implemented in a one digital tool by using jupyter notebooks. In order to achieve the goal of our seminar, we demonstrated various ways of using digital tools in the seminar. For example, we demonstrated with practical examples and applications that jupyter notebooks can be used to administer a multiple-choice test, homework, project and learning cards and even peer assessment.

As feedback is part of formative assessment and one of the most important approaches to improve the teaching and learning process, tools were introduced to show that it is possible to give quick digital feedback to students. These digital tools have many advantages for fast and effective feedback, especially in large and heterogeneous classrooms (e.g., reducing time, providing manual or automatic feedback, quickly sharing reports of results with students). We also provided that not only existing tools but also new applications can be developed with the many programming languages provided by jupyter. For example, we indicated that student teachers from history or language programs who choose to attend the seminar were able to gain basic programming skills. In addition, we also presented experiences with the advantages and benefits of the jupyter notebooks in terms of creating and sharing learning and teaching materials and the visual superiority of the jupyter notebooks.

One of the main objectives of the seminar was to enable student teachers to experience differential assessment activities based on open access resources and to develop positive attitudes and belief in their self-efficacy skills, despite the learning challenges that accompany the process. We designed simple content and activities for the assessment that would be appropriate for the different teaching levels at which the student teachers would be teaching. We also ensured that the assessment content and activities would include practical advantages over paper-based practices, such as examples for animating the question cards, especially for student teachers who would be teaching at lower levels of education. These contents made assessment activities interesting for student teachers. Furthermore, fast and effective feedback is one of the main advantages of digital assessment tools. As an alternative to paper-based applications, we designed notebooks that provides fast and effective feedback in a short time. For example, instead of spending a class hour on assessment, they experienced the ease of doing it in a short time and presenting the results quickly in visual form and tables. In addition, the markdown content was created in a similar format to the real content used at different educational levels. For example, student mathematics teachers experienced that they were able to prepare and write markdowns with mathematical formulas and visuals in a very short time.

In the seminar, the differences in the profiles of the participants, the subject and the level of instruction affect the attitudes. In addition, in experimental studies, participants' prior knowledge also plays a crucial role. For example, a difficult course content may not be challenging for student teachers who have prior knowledge, while it may be disappointing for students who do not have prior knowledge. In addition, when the content is interesting, it can also have an impact on positive attitudes. Furthermore, the practical advantages of designing and using the jupyter notebooks need to be directly experienced. For example, the student teacher might ask themselves: Why should I learn and use a digital application instead of a traditional paper and pencil application? For this reason, the notebooks were designed to provide simple rather than complex course content, with examples that can be adapted to different levels of education. Furthermore, we introduced applications (e.g. visual advantages, time saving, etc.) that demonstrated the benefits that notebooks can provide. After practicing the different types of examples, the student teachers designed the assessment activities for their preferred subject, course content and instructional level. In the pre-tests, all the student teachers did not have prior knowledge about the use of jupyter notebooks. However, even the students with specific knowledge of a programming language were able to use it to design their project assignments and apply it to their preferred content.

The main emphasis of the formative assessment approach is the identification of learning gaps in a timely manner, providing quick feedback and making appropriate corrections. Although designing formative assessment activities with a digital approach has many advantages over traditional paper-and-pencil applications, it is important to emphasize the challenges that can accompany the process. First of these challenges is the necessity of appropriate technical conditions. Another challenge is the unfamiliarity of the student teachers with an application form. For example, although open access resources are very popular, an existing prejudice against coding or complex applications has accompanied the process. In addition, the documentation, often in English, can be challenging for local language practitioners. In this work, we have encouraged the use of artificial intelligence supports, especially to address the difficulties in writing codes. We also scheduled regular office hours to address preconceptions and learning disabilities related to programming. Providing additional support when they needed it played a role in changing existing negative attitudes. In addition, with quick feedback, practitioners found quick and practical solutions to mistakes. During the implementation of the jupyter notebooks, it was again important to have all participants on the same server and to share the notebooks quickly and securely. In addition, more tools and applications should be developed for a variety of assessment approaches. For example, instead of having to write code, practitioners should be able to quickly integrate one of the many available applications into the notebooks or integrate and adapt existing applications into the course content in the way they prefer.

Concerning the achievement of the objectives of our seminar, it was observed that even student teachers with very minimal programming skills and a background in digital applications quickly adapted to the process with the jupyter notebooks. Moreover, the end-of-semester project assignments prepared by the student teachers showed that all of them were successful in producing a digital notebook with formative assessment items and feedback. In addition to this way of evaluating of aim achievement of the seminar, a pre-posttest was also implemented in the seminar. The results supported the assumption that student teachers had a change in their attitudes: For instance, in the posttest positive attitudes increased significantly compared to the pre-test results. Furthermore, student teachers' self-efficacy beliefs changed significantly from pre- to posttest and on average they perceived themselves perceived themselves as more self-efficacious. This is consistent with other studies that implemented professional development or training courses for promoting teachers' competences regarding digital tools (e.g. Reinhold et al., 2021; Thurm & Barzel, 2020) and especially jupyter notebooks (Nwulu et al., 2021; Bascuñana et al., 2023).

It should also be emphasized that jupyter notebooks have been used as a teaching and learning tools in many fields (e.g., Cardoso et al., 2019; Reades, 2020), thanks to their numerous practical application advantages (e.g., excellent visualization and simulation features). In this study, it is emphasized that not only learning and teaching processes but also assessment activities can be done by using jupyter notebooks. Jupyter notebooks can be used by different researchers as learning and teaching tools in different disciplines or adapted for different assessment approaches. In addition, different course content can be prepared using jupyter notebooks and their effectiveness in learning and teaching processes can be tested. Particularly in digital learning and teaching processes and assessment situations, the practical benefits that jupyter notebooks can offer need to be explored.

## 5 Limitations

In this study there were various limitations, which were described below. One limitation is that the student teachers who participated in the seminar came from different disciplines and had different prior skills and beliefs. The procedure was also accompanied by the student teachers' adaptation to existing learning and teaching methods and practices. It should also be noted that even if our findings are consistent with previous research, generalizability is limited due to the small sample size and the fact that the study is based on an experimental approach. Consequently, similar studies could be conducted in different fields to confirm and extend the findings of this study. In particular, these approaches could be replicated using bigger sample groups and for more prolonged periods of time. Furthermore, assessment tools such as nbgrader, which can be quickly integrated into jupyter notebooks, could be developed for specific formative approaches. In addition, more qualitative approaches such as interviews should be used to investigate the impact of training courses for student teachers, in order to gain a deeper insight into the impact of the courses from the participants' perspective.

## 6 Conclusion

It was observed that despite the challenges of implementing the seminar with a heterogeneous group of student teachers, the seminar was partially successful in achieving its objectives of introducing student teachers to digital tools, transferring skills and developing positive beliefs. Student teachers lacking prior knowledge about programming, especially in the first weeks of the seminar, and had concerns or prejudices about whether digital applications could be programmed directly, emphasize the need for further research on the challenges of implementing new digital approaches in teaching and learning processes. Following suggestions can be made based on the applications in this study. For example, applications such as mind maps and jupytercards attracted the interest of student teachers. In contrast to paper and pencil applications, jupyter notebooks emphasize the understanding and comprehension of the differences between concepts thanks to the visual advantages they provide. In addition, for different assessment approaches, some tools such as nbgrader can be developed to facilitate quick and practical feedback. The numerous advantages of the jupyter notebooks could thus become more attractive to practitioners with less good programming skills.

**Acknowledgements** We would like to thank Digital *and Data Literacy in Teaching Lab* Team for the suggestions.

**Author contributions** GYT developed the original idea and study design, conducted the data analyses, wrote the codes and created the notebooks. JB contributed to the design of the formative assessment approaches and the development of the scales, as well as the writing of the introduction and discussion section of the paper. TP contributed to the creation of the jupyter notebooks, writing the codes and writing the introduction to the article. Both GYT and JB conducted the seminar and TP assisted the students in some sessions.

**Funding** Open Access funding enabled and organized by Projekt DEAL."This study was funded by the" Digital *and Data Literacy in Teaching Lab* (DDLitLab) of the University of Hamburg with funds from the *Foundation for Innovation in Higher Education".*

**Data availability** All codes, notebooks and data sets used in the present study are available from the corresponding author on reasonable request.

## Declarations

**Competing interests** The author declares that they have no competing interests**.**



## References

Al-Gahmi, A., Zhang, Y., & Valle, H. (2022). Jupyter in the Classroom: An Experience Report. In Proceedings of the 53rd ACM Technical Symposium on Computer Science Education - Volume 1 (SIGCSE 2022) (Vol. 1, pp. 425–431). Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3478431.3499379

Amoudi, G., & Tbaishat, D. (2023). Interactive notebooks for achieving learning outcomes in a graduate course: a pedagogical approach. *Education and Information Technologies*, *28*(12), 16669–16704.

Andersson, C., & Palm, T. (2017). Characteristics of improved formative assessment practice. *Education Inquiry*, *8*(2), 104–122.

Balta, N., & Tzafilkou, K. (2019). Using Socrative software for instant formative feedback in physics courses. *Education and Information Technologies, 24*(1), 307–323. https://doi.org/10.1007/s10639-018-9773-8

Balta, N., Perera-Rodríguez, V. H., & Hervás-Gómez, C. (2018). Using Socrative as an online homework platform to increase students' exam scores. *Education and Information Technologies, 23*, 837–850. https://doi.org/10.1007/s10639-017-9638-6

Bandura, A., & Wessels, S. (1997). *Self-efficacy* (pp. 4–6). Cambridge University Press.

Barba, L. A., Barker, L. J., Blank, D. S., Brown, J., Downey, A. B., George, T., ... & Zingale, M. (2019). Teaching and learning with Jupyter. Retrieved March 30, 2025, from https://jupyter4edu.github.io/jupyter-edu-book

Bascuñana, J., León, S., González-Miquel, M., González, E. J., & Ramírez, J. (2023). Impact of Jupyter Notebook as a tool to enhance the learning process in chemical engineering modules. *Education for Chemical Engineers*, *44*, 155–163.

Baszuk, P. A., & Heath, M. L. (2020). Using Kahoot! to increase exam scores and engagement. *Journal of Education for Business, 95*(8), 548–552. https://doi.org/10.1080/08832323.2019.1707752

Baumert, J., & Kunter, M. (2013). The COACTIV model of teachers' professional competence. *Cognitive activation in the mathematics classroom and professional competence of teachers: Results from the COACTIV project* (pp. 25–48). Springer, US.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5–25.

Bennett, R. E. (2015). The changing nature of educational assessment. *Review of Research in Education, 39*(1), 370–407.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Børte, K., & Lillejord, S. (2024). Learning to teach: Aligning pedagogy and technology in a learning design tool. *Teaching and Teacher Education*, *148*, 104693.

Bridgeman, B. (2009) Experiences from large-scale computer-based testing in the USA. In F. Scheuermann & J. Björnsson (Eds.), *The Transition to Computer-Based Assessment*. (Luxembourg, Office for Official Publications of the European Communities).

Cardoso, A., Leitão, J., Teixeira, C. (2019). Using the Jupyter Notebook as a Tool to Support the Teaching and Learning Processes in Engineering Courses. In: Auer, M., Tsiatsos, T. (eds) The Challenges of the Digital Transformation in Education. ICL 2018. Advances in Intelligent Systems and Computing, vol 917. Springer, Cham. https://doi.org/10.1007/978-3-030-11935-5_22

Chowdhury, S. A., Khalid, M. S., & Arefin, A. S. M. S. (2021). Digital Learning Technology Blend in Assessment Activities of Higher Education: A Systematic Review. In M. A. Impedovo, M. S. Khalid, K. Kinley, & M. C. K. Yok (Eds.), *Blended Learning in Higher Education* (pp. 102–111). Aalborg Universitetsforlag.

Chua, Y. P. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, *28*(5), 1580–1586.

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481–509.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. Routledge.

Ferdig, R. E., Baumgartner, E., Hartshorne, R., Kaplan-Rakowski, R., & Mouza, C. (Eds.). (2020). *Teaching, technology, and teacher education during the COVID-19 pandemic: Stories from the field*. Association for the Advancement of Computing in Education.

Fuster-Guilló, A., Pertegal-Felices, M. L., Jimeno-Morenilla, A., Azorín-López, J., Rico-Soliveres, M. L., & Restrepo-Calle, F. (2019). Evaluating impact on motivation and academic performance of a game-based learning experience using Kahoot. *Frontiers in Psychology, 10*, 2843. https://doi.org/10.3389/fpsyg.2019.02843

Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915–945. https://doi.org/10.3102/00028312038004915

Garrett, N., Thoms, B., Alrushiedat, N., & Ryan, T. (2009). Social ePortfolios as the new course management system. *On the Horizon, 17*(3), 197–207.

Gurung, M. P., Paudel, K., & Kadel, R. (2020, March). Comparative analysis of popular digital assessment tools for formative assessment on networking courses. In *INTED2020 Proceedings: 14th International Technology, Education and Development Conference* (pp. 3114–3122). IATED. https://doi.org/10.21125/inted.2020.0775

Han, I., Shin, W. S., & Ko, Y. (2017). The effect of student teaching experience and teacher beliefs on pre-service teachers' self-efficacy and intention to use technology in teaching. *Teachers and Teaching*, *23*(7), 829–842.

Hanč, J., Štrauch, P., Paňková, E., & Hančová, M. (2020). Teachers' perception of Jupyter and R Shiny as digital tools for open education and science (arXiv:2007.11262) . arXiv. https://doi.org/10.48550/arXiv.2007.11262

Harlen, W., & James, M. (1997). Assessment and Learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice, 4*(3), 365–379. https://doi.org/10.1080/0969594970040304

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

Howells, K. (2018). The future of education and skills: education 2030: the future we want. Paris OECD. http://www.oecd.org/education/2030/E2030%20Position%20Paper%20(05.04.2018).pdf

Howson, O., (2023). Leveraging Jupyter notebooks in assessment development, completion and marking to reduce cognitive load and minimise errors [paper presentation]. I-HE2023. Zenodo. https://zenodo.org/records/10123535

Hsu, P. S. (2013). Examining changes of preservice teachers' beliefs about technology integration during student teaching. *Journal of Technology and Teacher Education, 21*(1), 27–48.

Jupyter, et al. (2019). nbgrader: A tool for creating and grading assignments in the jupyter notebook. *Journal of Open Source Education, 2*(11), 32. https://doi.org/10.21105/jose.00032

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice, 30*(28–37), 19. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., ... & Willing, C. (2016). Jupyter Notebooks–a publishing format for reproducible computational workflows. In *Positioning and power in academic publishing: Players, agents and agendas* (pp. 87-90). IOS press.

Kohnke, L., & Moorhouse, B. L. (2022). Using Kahoot! to gamify learning in the language classroom. *RELC Journal, 53*(3), 769–775. https://doi.org/10.1177/00336882211040270

Lee, H. Y., Wu, T. T., Lin, C. J., Wang, W. S., & Huang, Y. M. (2024). Integrating Computational Thinking into Scaffolding Learning: An Innovative Approach to Enhance Science, Technology, Engineering, and Mathematics Hands-On Learning. *Journal of Educational Computing Research, 62*(2), 431–467. https://doi.org/10.1177/07356331231211916

Liu, S. H. (2011). Factors related to pedagogical beliefs of teachers and technology integration. *Computers & Education, 56*, 1012–1022.

Liu, S.-H. (2012). A multivariate model of factors influencing technology use by preservice teachers during practice teaching. *Educational Technology & Society, 15*, 137–149.

Nwulu, N. I., Damisa, U., & Gbadamosi, S. L. (2021). Students Perception about the Use of Jupyter Notebook in Power Systems Education. *International Journal of Engineering Pedagogy (iJEP), 11*(1), 78–86. https://doi.org/10.3991/ijep.v11i1.14769

OECD (2005). Formative assessment: Improving learning in secondary classrooms. Retrieved March 30, 2025, from http://www.oecd.org/publications/Policybriefs.

Palm, T., Andersson, C., Boström, E., & Vingsle, C. (2017). A review of the impact of formative assessment on student achievement in mathematics. *Nordic Studies in Mathematics Education, 22*(3), 25–50.

Petersohn, M., Schöbel, K., & Thor, A. (2023). Kopplung von Jupyter Notebooks mit externen E-Assessment-Systemenam Beispiel des Data Management Testers. In *E-Assessment und Feedback: 21. Fachtagung Bildungstechnologien (DELFI)* (pp. 85–90). Gesellschaft für Informatik e.V. https://doi.org/10.18420/delfi2023-16

Posnanski, T. J. (2002). Professional development programs for elementary science teachers: An analysis of teacher self-efficacy beliefs and a professional development model. *Journal of Science Teacher Education, 13*(3), 189–220.

Praetorius, A. K., & Charalambous, C. Y. (2018). Classroom observation frameworks for studying instructional quality: Looking back and looking forward. *ZDM, 50*(3), 535–553. https://doi.org/10.1007/s11858-018-0946-0

Price, M., Handley, K., Millar, J., & O'donovan, B. (2010). Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education, 35*(3), 277–289.

Reades, J. (2020). Teaching on Jupyter - Using notebooks to accelerate learning and curriculum development. *Region: The Journal of ERSA, 7*(1), 21–34. https://doi.org/10.18335/region.v7i1.282

Reinhold, F., Strohmaier, A., Finger-Collazos, Z., & Reiss, K. (2021). Considering teachers' beliefs, motivation, and emotions regarding teaching mathematics with digital tools: The effect of an in-service teacher training. *Frontiers in Education*, *6*, 723869. https://doi.org/10.3389/feduc.2021.723869

Rodrigues, F., & Oliveira, P. (2014). A system for formative assessment and monitoring of students' progress. *Computers & Education*, *76*, 30–41.

See, B. H., Gorard, S., Lu, B., Dong, L., & Siddiqui, N. (2021). Is technology always helpful? A critical review of the impact on learning outcomes of education technology in supporting formative assessment in schools. *Research Papers in Education, 37*(6), 1064–1096.

Shea, J. (2021). JupyterQuiz (Version 1.5) [Computer software]. Retrieved March 30, 2025, from https://github.com/jmshea/jupyterquiz

Shute, V., & Rahimi, S. (2017). Review of computer-based assessment for learning in elementary and secondary education. *Journal of Computer Assisted Learning, 33*(1), 1–19. https://doi.org/10.1111/jcal.12172

Shute, V. J., Hansen, E. G., & Almond, R. G. (2008). You can't fatten a hog by weighing it–or can you? evaluating an assessment for learning system called ACED. *International Journal of Artificial Intelligence in Education*, *18*(4), 289–316.

Speer, A., & Eichler, A. (2022). Developing prospective teachers' beliefs about digital tools and digital feedback. *Mathematics, 10*(13), 2192. https://doi.org/10.3390/math10132192

Sullivan, P., McBrayer, J. S., Miller, S., & Fallon, K. (2021). An examination of the use of computer-based formative assessments. *Computers & Education, 173*, 104274. https://doi.org/10.1016/j.compedu.2021.104274

Thomson, C., & Dawes, S. (2022). Survey on student experiences with Jupyter Notebooks on TM351: Detailed analysis. Open University. https://www5.open.ac.uk/scholarship-andinnovation/esteem/sites/www.open.ac.uk.scholarship-andinnovation.esteem/files/resources/Dawes%20and%20Thomson,%20Jupyter%20Notebooks%20detailed%20analysis.pdf

Thurm, D., & Barzel, B. (2020). Effects of a professional development program for teaching mathematics with technology on teachers' beliefs, self-efficacy, and practices. *ZDM Mathematics Education*, *52*(7), 1411–1422. https://doi.org/10.1007/s11858-020-01158-6

Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction, 15*(6), 589–602.

Voss, T., Kleickmann, T., Kunter, M., & Hachfeld, A. (2013). Mathematics Teachers' Beliefs. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Cognitive activation in the mathematics classroom and professional competence of teachers. Results from the COACTIV Project* (pp. 249–271). Springer.

Wang, H., & Shin, C. D. (2010). Comparability of computerized adaptive and paper-pencil tests. *Test, Measurement and Research Service Bulletin*, *13*, 1–7.

Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education, 149*, 103818. https://doi.org/10.1016/j.compedu.2020.103818

Wang, A. I., Elvemo, A. A., & Gamnes, V. (2014). Three social classroom applications to improve student attitudes. *Education Research International*, *2014*(1), 259128.

Wash, P. D. (2014). Taking advantage of mobile devices: Using Socrative in the classroom. *Journal of Teaching and Learning with Technology, 3*(1), 99–101. https://doi.org/10.14434/jotlt.v3n1.5016

Wiliam, D. (2010). An integrative summary of the research 1 literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18–40). Routledge.

Wiliam, D. (2010b). The role of formative assessment in effective learning environments. The nature of learning: Using research to inspire practice, 135–155.

Wiliam, D. & Thompson, M. (2008). Integrating assessment with learning: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Routledge: Taylor & Francis Group. https://discovery.ucl.ac.uk/id/eprint/10001162/

Yuster, D. (2022). *Using Python Jupyter Notebooks + nbgrader as a Homework System*. College of New Jersey.

Zastre, M. (2019). Jupyter notebook in CS1: An experience report. In *Proceedings of the Western Canadian Conference on Computing Education* (pp. 1–6). Association for Computing Machinery. https://doi.org/10.1145/3314994.3325072

Zhyhadlo, O. Y. (2022). Application of digital game-based tools for formative assessment at foreign language lessons. *Information Technologies and Learning Tools, 87*(1), 139–150. https://doi.org/10.33407/itlt.v87i1.4703