The European Educational Researcher

http://www.eu-er.com

SELFIEfor(STUDENT)TEACHERS? A Mixed-Methods Study on the Suitability and Validity of the Self-Reflection Tool for Student Teachers

Christian Seyferth-Zapf

University of Bayreuth, Germany

Cindy Bärnreuther & Melanie Stephan

Friedrich-Alexander-University Erlangen-Nuremberg, Germany

Matthias Ehmann & Maria Seyferth-Zapf

University of Bayreuth, Germany

Abstract: This paper presents a study focusing on the feasibility and validity of the SELFIEforTEACHERS self-assessment tool for student teachers, addressing three research questions. Using a mixed methods approach with a convergent parallel design. The study provides insights into quantitative and qualitative aspects of the self-assessment of student teachers' digital competences. Firstly, it is analyzed how student teachers assess their own media-related digital competences using SELFIEforTEACHERS. Secondly, uttered thought processes from student teachers were examined when using the self-assessment tool. Finally, the extent to which SELFIEforTEACHERS is suitable for the self-assessment of digital competences of student teachers was considered. The study was conducted at two German universities offering teacher training programmes to ensure a comprehensive survey of the target group. In the quantitative part, the SELFIEforTEACHERS instrument was used, comprising 32 items reflecting different digital competences based on the DigCompEdu framework. An overall sample of n = 127 student teachers who had completed more than three semesters of their studies was surveyed. Additionally, thoughts about the instrument were collected from the participants using the thinking aloud method (n_{think} = 9). The data analysis revealed that, on average, student teachers assign themselves at competence level B1. However, the validity of this assessment raises questions, particularly regarding the assumption of equidistant competence levels and their relevance for student teachers. The qualitative results revealed that student teachers encountered challenges using the self-assessment tool, including problems with the terminology used and the hierarchical structure of the competence levels. The integration of qualitative and quantitative data provided partial validation of the results and enabled a comprehensive discussion of the suitability of SELFIEforTEACHERS for student teachers' self-assessment of digital competences. The study concludes with recommendations for refining the self-assessment tool to better meet the needs of student teachers and increase its validity in the context of teacher education.

Keywords: Teacher Education; Student Teachers; Digital Competences; DigCompEdu; SELFIEforTEACHERS
DOI: https://doi.org/10.31757/euer.732

Introduction

Digital media offers various opportunities and challenges, especially for students and teachers. The COVID-19 pandemic and all the peculiarities that have come along with it regarding face-to-face teaching and remote learning emphasised the significance of digital media. However, the mere use of digital media itself does not guarantee successful teaching and learning as it heavily depends on the teacher's digital competence (beliefes) (Guggemos & Seufert, 2021; Quast, Rubach & Lazarides, 2021; Scheiter, 2021). Against this background, the pandemic was considered as amplifier for media-related digital competences among German teachers because they have assessed their competences of using digital media in class more positively since then (Endberg & Lorenz, 2022; Lorenz et al., 2022). In times before the pandemic, German teachers ranked below-average compared to their international colleagues not only in terms of using digital media for teaching and learning but also in fostering ICT-related

competences among their students (Drossel et al., 2019). From a theoretical point of view there is a well-established international and national discourse about media-related digital competences among teachers leading to numerous models and conceptualizations such as TPACK (Mishra & Koehler, 2006), DigCompEdu (Redecker, 2017), the model of *Medienpädagogische Kompetenz* (Blömeke, 2000; Herzig & Martin, 2018; Tulodziecki, 2021), or the K19-framework (DCB, 2017) and many more. Fortunately, many of these models were operationalized in the past years resulting in a decent amount of predominantly self-assessment tools (Economou, 2023a; Ghomi & Redecker, 2019; Herzig et al., 2015; Quast, Rubach & Porsch, 2023; Sailer et al., 2021; Schmid, Brianza & Petko, 2020; Schmidt et al., 2009; Vejvoda et al., 2024). Apart from that, a closer look at these instruments reveals that although some of the above-mentioned studies include student or pre-service teachers in their samples none of them – except for the M^3K (Herzig et al. 2015), which will be discussed later – was explicitly designed for student teachers. Surveys of student teachers and teachers are always conducted together and not separately. This also applies to the SELFIEforTEACHERS (Economou, 2023a). A self-assessment tool for teachers based on the DigCompEdu framework clearly emphasizing in-service teaching practices (e.g. item example: *I have tried using digital technologies to support and/or enhance my teaching practice*) on the one hand but on the other hand it also wants to addresses teacher educators aiming "to help their students design their learning pathways to further develop their digital competence or plan a course for their students." (Economou, 2023b, p. 2). In this sense, the study presented in this paper focuses on the feasibility and validity of SELFIEforTEACHERS for student teachers while providing answers to the following research questions:

RQ 1: How do student teachers rate themselves regarding their own media-related digital competences using the SELFIEforTEACHERS?

RQ 2: What thought processes are uttered by student teachers while working on the SELFIEforTEACHERS?

RQ 3: To what extent is the SELFIEforTEACHERS suitable for the self-assessment of media-related and digital competences in student teachers?

Theoretical background and current state of research

In a systematic study of research about teachers' digital competence beliefs, Rubach and Lazaridis (2023) conclude that there is a "great variety of different frameworks [which] leads to highly diverse definitions and categorizations of teachers' competence" (p. 192) implying a lack of comparability. Nevertheless, it is inevitable to provide a brief overview of digital competence models relevant to student teachers. Criteria for model inclusion were: 1.) significance in the national and international discourse about media related competences among (student) teachers, 2.) transfer to empirically tested and validated instruments and 3.) suitability and applicability for student teachers. The selected models serve as contextual framework and support the argumentation in favour of the SELFIEforTEACHERS as instrument for answering the above-mentioned research questions.

Digital competence models (for student teachers)

In the following, the contextual framework for conducting this study with SELFIEforTEACHERS is set up. Various theoretical and didactic models for skills acquisition are presented, as well as the framework of the European Union.

Modelling and measuring media pedagogical competence (M3K)

From a national point of view, the discourse about *Medienpädagogische Kompetenz* is probably the most established one since its roots reach back into the 1990s. Deriving from debates about media-related competence standards for initial teacher education, the concept of *Medienpädagogische Kompetenz* is supposed to hold true for all phases of teacher education although it clearly emphasises the first phase (Tulodziecki, 2012). Over the years, different approaches of *Medienpädagogische Kompetenz* (e.g. Blömeke, 2000; Gysbers, 2008; Herzig et al., 2016; Tulodziecki, 2012) have developed leading to a fundamental understanding of the construct as an interplay of competence areas which Herzig et al. (2016) summarise in their M^3K-model as: 1. using digital media for teaching and learing (*Mediendidaktik*), 2. applying media literacy education (*Medienerziehung und -bildung),* 3. media-related school development (*medienbezogene Schulentwicklung*). In addition, media-related attitudes and self-efficacy beliefs as well as media-technological knowledge are considered as favourable prerequisites for *Medienpädaogische Kompetenz* (Herzig & Martin, 2018). The M^3K was transferred into a competence test containing subscales with scenario-based items for each competence area and additional five scales for competence-related attitudes, media-related self-efficacy beliefs and media-technological knowledge with a total of 102 items (ibid.). Unfortunately, results of a validation study with student teachers from seven universities in Germany (n = 919) revealed serious problems concerning the internal consistency of the subscales (ibid.), leading to the fact that "no finalized and validated instrument has been published so far" (Tiede, 2019, 111).

Instrument for core competences (IN.K19+)

Another national model that was transferred into a self-assessment instrument for teachers is the K19+ competence framework developed by the *Digital Campus of Bavaria* research group (DCB, 2017). It describes 19 media- and technology-related skills for teachers (i.e. *assessing students' basic digital skills*) divided into categories such as *planning*, *implementing*, *evaluating* and *sharing* (Sailer et al., 2021). These categories are supposed to represent different phases of teaching with and about digital media leading to "a closer connection between technology-related teaching skills and actual technology-related classroom learning activities" (Sailer et al., 2021, p. 2). Based upon the K19 framework, the scenario-based self-assessment instrument IN.K19+ was developed "encompassing knowledge and action-oriented facets of technology related skills" (Vejvoda et al., 2023, p. 528). IN.K19+ was validated in two studies with samples of student and in-service teachers showing promising results in terms of validity, factor structure and internal consistency (Sailer et al., 2021; Vejvoda et al., 2023). In its final version the instrument describes 19 scenarios with 57 items and a 5-point Likert scale while participants receive graphical feedback illustrating their skill level for each scenario (Vejvoda et al., 2023).

Technological Pedagogical Content Knowledge (TPACK)

Compared to the afore mentioned models M^3K and K19, TPACK can be considered as a rather one-sided model as it emphasises different areas of (technological) knowledge instead of skills or competencies and does not include aspects of media literacy education as it "focuses on teaching with media" (Tiede, Hobbs & Grafe, 2015, p. 537). The TPACK-

model was introduced by Mishra and Koehler (2006) and consists of the major knowledge areas *technological knowledge* (TK), *pedagogical knowledge* (PK) and *content knowledge* (CK) as well as its overlapping sections *pedagogical content knowledge* (PCK), *technological content knowledge* (TCK), *technological pedagogical knowledge* (TPK) and *technological pedagogical content knowledge* (TPACK). According to Schmid et al. (2024), "TPACK is one of the most cited frameworks within the field of educational technology research" (p. 1). One of the biggest achievements of more than 15 years TPACK framework is its major contribution to international empirical research. Over the years, many self-assessment scales have been developed and statistically validated using pre-service and student teachers (e.g. Chai et al., 2013; Schmid et al., 2020; Schmidt et al., 2009; Valtonen et al., 2015; Yurdakul et. al., 2012). Thus, TPACK research is also on the rise in German speaking countries focusing on framework adaptions and developments (e.g. Döbeli Honegger, 2021; Huwer et al., 2019; Schmid & Petko, 2020) as well as large scale studies among German in-services teachers based on a modified version of the TPACK scale form Schmidt et al. (2009) (e.g. Endberg & Lorenz, 2017; 2022).

Digital Competence Framework for Educators (DigCompEdu) – SELFIEforTEACHERS

Digital Competence Framework for Educators (DigCompEdu) is a framework of the European Commission's Joint Research Centre (JRC) to "capture and describe […] educator specific digital competences" (Redecker, 2017, p. 9), targeting "educators at all levels of education" (ibid., p.13). Unlike TPACK, DigCompEdu does not only focus on knowledge, but also describes skills and attitudes as part of teacher's digital competences (Caena & Redecker, 2019). Moreover, the intention of DigCompEdu is to "be applicable across all subjects […] [emphasising] the pedagogical element" (Ghomi & Redecker, 2019, p. 542). In this context, digital competences of educators are composed out of 22 competences which are grouped together into six areas (Redecker, 2017): *Area 1 Professional Engagement (1.1 organisational communication, 1.2 professional collaboration, 1.3 reflective practise, 1.4 digital continuous development); Area 2 Digital Resources (2.1 selecting, 2.2 creating & modifying, 2.3. managing, protecting, sharing); Area 3 Teaching and Learning (3.1 teaching, 3.2 guidance, 3.3 collaborative learning, 3.4 self-regulated learning); Area 4 Assessment (4.1 assessment strategies, 4.2 analysing evidence, 4.3 feedback & planning); Area 5 Empowering Learners (5.1 differentiation & personalisation, 5.2 accessibility & inclusion, actively engaging learners); Area 6 Facilitating Learner's Digital Competence (6.1 information & media literacy, 6.2 communication, 6.3 content creation, 6.4 responsible use, 6.5 problem solving)*. A closer look at these competences reveals that DigCompEdu is not only about the use of digital media for teaching and learning, but also about media education aspects, which are explicitly anchored in area 6. In addition, the overall framework is conceptualised as a cumulative progression model to help teachers identify their strengths and deficits (ibid.). Against this background, Redecker (2017) uses the taxonomy of the Common European Framework of Reference for Languages (CEFR) and Bloom's revised taxonomy of educational objectives (Anderson & Krathwohl, 2001) as major references for the development and description of six DigCompEdu competence levels ranging from A1 to C2.

Based on the DigCompEdu model, the JRC developed the self-assessment tool Check-In in 2018, which was online available on EU-Survey until January 2022. The instrument consisted of 22 items (one item for each competence in

DigCompEdu) showing excellent internal consistency values (*Cronbach's alpha* .934) (Ghomi & Redecker, 2019). However, the favoured one-competence-one-item-structure led to the problem, that in some cases "a choice had to be made between different aspects crucial to a given competence" (Economou, 2023a, p. 17) (e.g. *2.3 managing, protecting, sharing*). Another controversial aspect of the Check-In tool is that in sometimes the transfer of the progression model seemed inappropriate leading to a partial deletion or blending of competence levels (Ghomi & Redecker, 2019). In sum, the tool displays five answer options per item corresponding roughly with the taxonomy pattern of DigCompEdu. Depending on the chosen answer participants could score from 0 to 4 points per item adding up for a total of max. 88 points (ibid.).

In 2022, the Check-In tool (DigCompEdu) was replaced by the SELFIEforTEACHERS. As opposed to its predecessor, the SELFIEforTEACHERS did not pursue the one-competence-one-item strategy. Instead, some "existing items were split to accommodate different aspects of the competence […] [or] new items were added to emphasise current needs" (Economou, 2023a, p. 20). This led to an increased item pool of 32 in total. Furthermore, it seemed as if the progression model of DigCompEdu was transferred more accurately into the new self-assessment tool leading to the implementation of six competence levels A1 – Awareness to C2 – Innovation (Economou, 2023b). Each item starts with an introductory statement followed by six proficiency statements (in accordance with the competence levels), from which participants are asked to select the one that fits best to their individual competence level (Economou, 2023a). Instead of using a Likert scale, the approach of choosing one statement for each item is favoured, because according to Economou (2023a), it corresponds perfectly to the assumption that "each [competence] level reflects a 'range' of more complex abilities in a continuum rather than an absolute number" (p. 20). Depending on the selected proficiency statement, 0 to 6 points could be achieved per item summing up for a total of max. 192 points (ibid.). After completing the test, SELFIEforTEACHERS provides a graphical overview of the individual self-reflection overall results and the results by area (Economou, 2023b).

For the study presented in this paper, the DigCompEdu model and the self-assessment tool SELFIEforTEACHERS were used. Compared to other frameworks (e.g. TPACK), DigCompEdu can be considered as holistic approach since it addresses aspects of teaching and learning with and about media on the one hand and focuses on educators of different school types, subjects, and levels on the other. Moreover, it is a well-established international model for describing essential digital competences among teachers with implications and effects on the national level as well (Kultusministerkonferenz, 2021). Above all, the fact that the tool delivers personalised feedback to the individual proficiency level and is freely available in a German version made it particularly appealing for this study.

Methodology and Research Design

In the present study, a mixed methods approach with a convergent parallel design (Creswell, 2014) was pursued to get a comprehensive understanding of how student teachers self-assess their media-related digital competence (quantitative part) and what considerations guide them when using the self-assessment tool (qualitative part). The study was carried out at two German universities (University of Bayreuth, Friedrich-Alexander-University Erlangen-

Nuremberg) that offer teacher training to ensure a precise survey of the target group. A non-probabilistic random sample of student teachers who had completed more than three semesters of their studies was used to collect the data (Döring & Bortz, 2016). This approach ensured that the participants already had a basic familiarity with the teaching profession and its responsibilities. Participation in the survey was voluntary. However, as the content and results were taken up in the further course of the seminar, participation was advised. In addition, the students were able to benefit from the final summary of their current level of competence. The student teachers were free to decide whether they wanted to answer the questionnaire individually or verbalize their thoughts when completing the qualitative survey.

Quantitative part

For conducting the study, the original SELFIEforTEACHERS self-assessment tool in its German version (free of charge available at: https://education.ec.europa.eu/selfie-for-teachers) was rebuilt one-to-one. This also includes the following details of the original: the action verbs in each task were highlighted in colour, help texts were integrated using a scroll-over function, examples were added in brackets and the individual feedback at the end was prepared graphically. Moreover, questions about age, sex, associated university, study progress, and previously attended courses about digital media were added. Immediately before the start of the SELFIEforTEACHERS instrument, student teachers were supposed to rank themselves according to their estimated digital competence level (A1 to C2). The questionnaire contains a 7-point Likert scale for each item. The six DigCompEdu competence levels from A1 to C2 and an additional seventh answer option "I am not aware of the competence." can be selected. They should repeat the same after having finished the SELFIEforTEACHERS. The survey itself was conducted with LimeSurvey, which was hosted locally. For data analysis SPSS and R were used.

Qualitative part

The qualitative part of the study used the thinking aloud method (Hofmann, 2017), based on the SELFIEforTEACHERS quantitative survey instrument. This method made it possible to record the thoughts and feelings of the students during the self-assessment. Data was collected through video recordings and interview protocols. The participants were encouraged to verbalize their thoughts openly while answering the questionnaire. A standardized procedure was ensured by a guideline developed by the research group. The audiovisual material was then converted into transcripts. Simple transcription rules were deliberately applied in the sense of a semantic-content transcription (Dresing & Pehl, 2018). The qualitative data analysis was carried out with MAXQDA 2022, using an inductive qualitative content analysis according to Kuckartz (2018). By means of this step-by-step procedure and mutual exchange, the category system was systematically developed and refined in the sense of Kuckartz (2018). The coding and recoding were carried out by three coders who were in close contact with each other. Consensual coding in alternating tandems achieved a high intercoder reliability of Cohens Kappa between $\varkappa = .86$ and $\varkappa = 1.00$, which confirms the reliability of the analysis.

Results

Sample

The survey period lasted from 18/04/2024 to 24/07/2024. 236 teacher students of both universities were invited to take part in the survey. 188 questionnaires were started in total, of which 130 were completed. The response and dropout rates are therefore 80% and 31% respectively. The average completion time of the 121 people who were not surveyed using the thinking aloud method was 29.27 minutes. However, there is one data set where the processing time was 1204 minutes. This may be due to a delay in pressing the send button in the digital survey. If this data set is not considered, the average processing time is 19.48 minutes. A reading time calculator was used to determine the reading time for the category introductions and the first two answer choices for all items. For a fast reader this is approximately 8.2 minutes. When analysing the completion times, there were three records with a completion time of less than 7.2 minutes, in which 'I am not aware of this competence' was not selected for any answer. There is a significant time gap of 30 seconds to the next data sets, which are continuously distributed along the time axis. Therefore, these three data sets were excluded from further analysis as they cannot be considered serious. The average processing time for the remaining 118 data records was 19.80 minutes, not considering the special case mentioned above. In terms of processing time, the average processing time per item of the respective subscale decreases continuously from subscale 3 *Teaching and Learning* onwards. It is still 38.47 seconds for subscale *2 Digital Resources* and falls to 22.31 seconds for subscale 6 *Facilitating Learners' Digital Competence*. Further descriptive data are listed in table 1.

The adjusted sample size as shown in table 1 is $n = 127$ with 61 participants from University of Bayreuth (UBT) and 65 participants from Friedrich-Alexander-University Erlangen-Nuremberg (FAU). Each academic year in Germany is divided into two semesters. On average, the student teachers were in their fifth semester ($M_{sem} = 5.06$, $SD_{sem} = 2.49$) and aged between 19 and 49 ($M_{age} = 23.32$, $SD_{age} = 5.19$). Overall, there was an unbalanced proportion of sexes with 91 female and 36 male participants. School type refers to the school level for which the student teachers are studying. The distribution of school types among the student teachers was as follows (in brackets the number *n* of respondents and the planned duration of the degree programme in semesters): grammar school ($n = 45$, 9 semesters), primary school ($n = 40$, 7 semesters), secondary school (RS) ($n = 32$, 7 semesters), secondary school (MS) ($n = 6$, 7 semesters) and vocational school ($n = 3$, 10 semesters).

Table 1

Descriptive data of the sample[*]

Group	**Sample size**		**Age**	**Semesters**	**Sex**		**School type**[**]				
	UBT	**FAU**			**female**	**male**	**gram**	**prim**	**sec (RS)**	**sec (MS)**	**voc**
Thinking aloud (*n = 9*)	4 44 %	5 56 %	*M* = 23.89 *SD* = 3.48	*M* = 7.44 *SD* = 2.60	4 44 %	5 56 %	4 44 %	1 11 %	3 33 %	1 11 %	- -
Non thinking aloud (*n = 118*)	57 48 %	60 51 %	*M* = 23.28 *SD* = 5.31	*M* = 4.87 *SD* = 2.40	87 74 %	31 26 %	41 35 %	39 33 %	29 25 %	5 4 %	3 3 %
Total (*n = 127*)	61 48 %	65 51 %	*M* = 23.32 *SD* = 5.19	*M* = 5.06 *SD* = 2.50	91 72 %	36 28 %	45 35 %	40 31 %	32 25 %	6 5 %	3 2 %

Note: *Differences in the total number of participants are because not all participants provided information on every question/item. In this case, one participant did not mark his/her university affiliation and another person did not specify his/her school type. Numerical values without further specification are absolute numbers of participants. The percentage values are relative proportions of participants in relation to the total number *n* in the respective row.

**School types: gram = grammar school (Gymnasium), prim = primary school (Grundschule), sec (RS) = secondary school (Realschule (RS)), sec (MS) = secondary school (Mittelschule (MS)), voc = vocational school (Berufsschule).

In the survey using the thinking aloud method (Hofmann, 2017), four student teachers from University of Bayreuth (UBT) and five student teachers from Friedrich-Alexander-University Erlangen-Nuremberg (FAU) took part. Their results were also included in the quantitative data. Of the participants, five were male and four females. Four were studying to become grammar schoolteachers, two were studying to become secondary school teachers (MS/RS) and one was studying to become a primary school teacher. There was great heterogeneity regarding the semesters studied. Numbers ranged from four to twelve. It should be noted that all nine interviewees had already attended at least one course with a media education focus at the time of the survey. The category system was developed, reviewed, and revised in a circular process. The following steps were essential:

- Deductively derived initial structuring based on the research questions
- First trial coding with inductive specification of the categories
- Coding by at least two independent persons
- Calculation of intercoder reliability and revision of the codes based on the results
- In-depth discursive revision of the category system streamlining by merging categories, clarification and structuring
- Recoding
- Finalization of the category system

Quantitative Data - Self-assessed media-related and digital competences of student teachers (RQ 1)

Before analyzing the quantitative data used to answer RQ 1, it is necessary to take a closer look at the types of data generated by the SELFIEforTEACHERS. According to Economou (2023a), participants can score 0 to 6 points for each item leading to a maximum sum score of 192 points. From a statistical point of view, this approach requires the existence of a metric scale depending on the equidistance between each proficiency statement. However, there is no theoretical or empirical evidence in SELFIEforTEACHERS supporting the equidistance assumption. Therefore, the following section provides median values (*Mdn*) in addition to the intended sum ($\sum$) and mean scores (*M*).

Descriptive Statistics

First, there was no evidence for a normal distribution of the sample values making it necessary to use non-parametric methods for data analysis. Overall internal consistency was excellent with .95, whereas *Cronbach's alpha* ranges at subscale level from .70 (subscale 2: *Digital Resources*) and .90 (subscale 6: *Facilitating Learner's Digital Competence*). In total, student teachers averaged a sum score of $\sum_{SELFIE} = 65.83$ ($SD_{\Sigma} = 23.45$), a mean score of $M_{SELFIE} = 2.53$ and a median of $Mdn_{SELFIE} = 2.00$ ($SD_{M/Mdn} = .81$). Table 2 provides an overview of mean and median values[1] at subscale level, showing the highest mean values for subscale 2 and the lowest for subscale 6.

Table 2

Mean and median values on subscale level

	Subscale 1 *Professional Engagement*	**Subscale 2** *Digital Resources*	**Subscale 3** *Teaching and Learning*	**Subscale 4** *Assessment*	**Subscale 5** *Empowering Learners*	**Subscale 6** *Facilitating Learners' Digital Competence*
M	2.31	2.60	1.93	1.74	1.74	1.72
SD_M	.79	.73	.87	.90	.88	.97
Mdn	2.00	2.00	2.00	1.00	2.00	1.50
SD_{Mdn}	1.00	.82	1.06	1.00	1.00	1.10

In addition to the results from SELFIEforTEACHERS, students were asked to rank themselves on the A1 to C2 competence levels before (t1) and after (t2) self-assessment. Surprisingly, no significant difference was found between Mdn_{SELFIE}, Mdn_{t1} and Mdn_{t2}, while the Friedman-Test revealed a significant difference when using the mean (M_{t1} = 2,35, M_{t2} = 2,17), χ^2 (2, n = 127) = 31.95, $p < .001$[2]. Two conclusions can be drawn from these results: first, student

[1] At subscale level, only mean and median values are reported. Sum scores are avoided, as each subscale consists of a different number of items, which implies a lack of comparability.

[2] For calculating M_{SELFIE} the mean from the individual competence levels (min. 1 = A1; max. 6 = C2) was used due to its comparability with M_{t1} and M_{t2} which were equally determined. For Mdn_{SELFIE} the median from the individual proficiency levels was used. In the further course, the mean value is calculated as the average of the individual responses to each item. Likewise median values are calculated of the item specific individual answers.

teachers tend to underestimate their media-related digital competences and second, the use of the instrument itself may cause a change in the self-assessment of the participants' competences. Moreover, results indicate that there is no statistically significant difference for the overall scale between the participants of the thinking aloud group (n_{think} = 9) and the rest (n_{rest} = 118), regardless of data type (U_{Σ} / U_M = 369.50, p = .13; U_{Mdn} = 413.50, p = .25). However, for subscale 2 (*Digital Resources*) significant differences for these two groups can be identified, when calculating with mean values resp. sum scores (U_{Σ} / U_M = 264.00, p = .012; U_{Mdn} = 313.00, p = .090).

Table 3 shows the distribution of competence levels (sum score) among the student teachers depending on their school type. The bulk of students ranked between A2 and B1, whereas A1, B2 and C1 were only achieved by a clear minority of participants. None of them reached proficiency level C2. Regarding differences in results between student teachers at different school types, it can be stated that Kruskal-Wallis-Test did not show any significant inequalities at first sight ($H_{MW/\Sigma}$ (4, n = 126) = 6.71, p = .15; H_{Mdn} (4, n = 126) = 6.01, p = .20). However, a pairwise comparison of the different school types reveals significant differences in overall results for grammar and primary school (U_{Σ} / U_M = 648.00, p = .026; U_{Mdn} = 673.00, p = .033) A closer look at subscale level proves that that inequalities between grammar and primary school are particularly large ($d_{MW/\Sigma}$ = .63, d_{Mdn} = .53) in subscale 6 (*Facilitating Learner's Digital Competence*). This goes hand in hand with the observation that student teachers for primary school chose answer option "I am not aware of this competence" (0 points) for item 6.1 *Information and data literacy* six times more often than their colleagues from grammar school (see Bärnreuther, 2024 on the relevance of digital competence (non)-awareness in SELFIEforTEACHERS among student teachers). These differences may be due to the assumption, that it might be more difficult for student teachers at primary schools to prepare learning activities that "require students to critically search, evaluate and manage information and data from different digital environments" (Economou, 2023a, p. 76).

Table 3

*Distribution of proficiency levels depending on the school type / school level for which the student teachers are studying**

	A1 (0-32 pts)	**A2 (33-64 pts)**	**B1 (65-96 pts)**	**B2 (97-128 pts)**	**C1 (129-160 pts)**	**C2 (161-192 pts)**	**sum**
gram	1	19	22	3	-	-	45
prim	3	25	10	1	1	-	40
sec (RS)	2	16	9	3	2	-	32
sec (MS)	-	2	3	1	-	-	6
voc	-	2	-	-	1	-	3
total	**6**	**64**	**44**	**8**	**4**	**-**	**126**

Note: *School types: gram = grammar school (Gymnasium), prim = primary school (Grundschule), sec (RS) = secondary school (Realschule (RS)), sec (MS) = secondary school (Mittelschule (MS)), voc = vocational school (Berufsschule). One person did not specify his/her school type.

Description of the correlations

Another observation was that overall results from teacher students who had taken part in courses on aspects such as teaching and learning with and about digital media rated themselves (significantly) better than those who had not taken a course in this field yet (U_{Σ}/ U_M = 1315.50, p = .01; U_{Mdn} = 1486.50, p = .09). This is particularly evident in subscale 2 *Digital Resources*, where a large effect ($d_{MW/\Sigma}$= .70, d_{Mdn} = .51) can be described between both groups. As already mentioned, the SELFIEforTEACHERS is based on the assumption of equidistant data. Statistically, this assumption is not tenable and it can only be assumed that the response options are ordinally scaled. Therefore, for consistency, correlations between the number of courses attended and the test results are considered for both cases (Pearson correlation for equidistantly scaled data, Spearman correlation for ordinally scaled data). There was a significant correlation for the number of courses attended (max. 5) and the overall results ($r_{Spearman}$ (127) = .179, p < .05; $r_{Pearson}$ (127) = .303, p < .05). On subscale level, correlations were found for the first three subscales. For subscale 5 a correlation was found when assuming equidistant data (table 4). In particular, the high correlation between the results of subscale 2 and the number of courses attended may explain the different results in that area between the thinking aloud group (n_{think} = 9) and the rest (n_{rest} = 118). While all participants in n_{think} had attended courses addressing digital competences, only 36.4 % of the rest had done so.

Table 4

Correlations – number of courses/results

	Subscale 1 *Professional engagement*	**Subscale 2** *Digital resources*	**Subscale3** *Teaching and learning*	**Subscale 4** *Assessment*	**Subscale 5** *Empower-ing learners*	**Subscale 6** *Facilitating learners' digital competence*	**Overall**
Mdn							
Spearman							
r	.253**	.269**	.234**	.097	.117	.048	.179*
p	.004	.002	.008	.276	.192	.588	.043
n	127	127	127	127	127	127	127
MW/Σ							
Pearson							
r	.328**	.354**	.322**	.173	.227*	.141	.303*
p	<.001	<.001	<.001	.052	.010	.113	<.001
n	127	127	127	127	127	127	127

Note: ** p < .001 (2-sided); * p < .05 (2-sided)

Conclusions and recommendations for further development based on the quantitative results

In response to RQ 1 it can be stated that, student teachers using the SELFIEforTEACHERS assign themselves to competence level B1 on average ($\sum_{SELFIE}$ = 65.83; $SD_{\sum}$ = 23.45). However, data results and the overall conception of the instrument raises questions to issues which need further elaboration:

1. There is neither empirical nor theoretical evidence for the assumption of equidistant competence levels, which makes the proposed way of generating a sum score from the individual points for each item highly questionable. This leads to an overrepresentation of those areas containing many items (e.g. Area 1) on the one hand and to different results, if the statistically more appropriate median is used instead (e.g. Mdn_{SELFIE} = 2.00 would imply a competence level of A2 on average).
2. The average competence level B1 is characterized by the action verb "use" (Economou, 2023a, p. 22) which implies the integration and creative use of digital media in teaching practices (Redecker, 2017). However, the underlying sample consisted of student teachers and their only teaching practice may either be a result of internships as part of their studies or from part-time jobs as temporary teaching assistants (which cannot be considered as norm). The fact that most participants rate themselves at this level suggests that the competence levels may have limited validity for the target group of student teachers.
3. The analyzed data revealed a significant positive correlation between the number of courses attended and the overall score in SELFIEforTEACHERS (table 4). This result corresponds to an understanding of competence development as cumulative process, which is in line with the general assumption of DigCompEdu as progression model (Redecker, 2017). Unfortunately, due to its statistical implementation (e.g. equidistant point ranges for each competence level, see table 2) SELFIEforTEACHERS suggests that the acquisition of competences is a linear process which does not correspond to reality.

In addition to the results from the quantitative data, qualitative data were collected and analyzed to capture thought processes from the student teachers which were expressed while working on the SELFIEforTEACHERS.

Qualitative Data - Thoughts of the students while working on the SELFIEforTEACHERS (RQ 2)

The finalized category system comprises ten main categories with further sub-categories (table 5). The main statements made by student teachers from seven categories are summarized below. The categories *Difficulties with demographic data (1)*, *General feedback on such questionnaires (9)* and aspects of *Non-awareness of digital competences (10)* are not specifically addressed (see Bärnreuther, 2024) for more information about digital competence (non)-awareness in SELFIEforTEACHERS). The interviews were conducted in German. The quotations for this article were translated into English by the authors.

Table 5

Category system – Content analysis (qual.)

Category	Sub-category	Number of codings
1. Difficulties with demographic data	*No sub-categories*	5
2. Answerability from student perspective	*2.1 Can be answered well*	9
	2.2 Difficult to answer	130
3. Terminology, specialized terms	*3.1 Comprehension problems*	35
	3.2 No comprehension problems	9
	3.3 Derive the term themselves	14
4. Answer formulation	*4.1 Positive*	34
	4.2 Negative	76
	4.3 Suggestions for improvement	22
5. Motivation in answering	*5.1 Positive*	10
	5.2 Negative	4
	5.3 Diminishing, answers are rather skimmed over	18
	5.4 Desire to drop out	3
6. Emotion in answering	*6.1 Positive*	36
	6.2 Negative	52
7. Formatting	*7.1 Positive*	15
	7.2 Negative	14
	7.3 Suggestions for improvement	11
8. Feedback on the evaluation	*8.1 Positive*	19
	8.2 Negative	17
	8.3 Suggestions for improvement	30
9. General feedback on such questionnaires	*No sub-categories*	8
10. Non-awareness of digital Competences	*10.1 Professional engagement*	31
	10.2 Digital resources	21
	10.3 Teaching and learning	23
	10.4 Assessment	20
	10.5 Empowering learners	17
	10.6 Facilitating learners' digital competence	18

Answerability from a student perspective

Some statements of the questionnaire are well suited to students because they have already dealt with relevant content during their studies: "You can look at it from the point of view of the media education seminar. If you then go in the direction of cyberbullying or fake news. But we definitely dealt with it actively." (I04_UBT: 125, also I02_FAU: 7; I04_UBT: 71, 91). On the one hand, it was stated much more frequently that it was considerably difficult to deal with the items from the student's perspective. Three students (or nine codes) included statements that confirmed suitability for students. However, it was unequivocally stated that there were clear indications in every interview that the students were struggling to find themselves in the answer options provided, with their current experiences (130 codes). For example, interview I04_UBT stated: "I would almost tend to turn it around, also in general, because we tend to analyze and evaluate in our studies and then the area of practical application is rather smaller." (I04_UBT: 71) Many answer options could only be answered on the condition that the students redefined them individually (e.g. colleagues as fellow students) (cf. I01_FAU: 21; I04_UBT: 54).

Terminology, specialized terms

Five students stated that they generally had no serious problems understanding the answers (I01_FAU; I02_UBT; I02_FAU; I03_UBT; I04_UBT). Three students showed uncertainty in understanding various terms. These included in particular the term “computational thinking” (I03_UBT; I04_UBT; I01_UBT; I05_UBT: 52), but also “creative commons license” (I01_UBT: 82; I05_UBT: 57), “digital artefacts” (I04_UBT), the description “professional learning activities" (I01_UBT: 64), curation (I01_UBT: 82), “external stakeholders” (I01_UBT: 92), debugging (I01_UBT: 170), “collaborative online activities” (I04_FAU: 12) and “formative and summative assessment” (I04_FAU: 37; I01_UBT: 106). The terms “tree structures”, “metadata” and “tags” also caused one interviewed student to ponder (I05_UBT: 65, 67). The student teachers interviewed gave an insight into how they tried to make sense of incomprehensible terms themselves with their statements: “Taking data management into account' [note: read from the questionnaire] I wouldn't know what you want from me at first reading. But perhaps the question opens to me when I answer the questions. (...) 'Data management methods' [note: read out from the questionnaire] I wouldn't know either. I'm being honest. (...) Okay. I assume GDPR [Author's note: General Data Protection Regulation] is some kind of data policy (...)”. (I01_UBT: 28). In some cases, the examples given proved to be helpful in deriving incomprehensible terms: “I probably do it via the examples now” (I04_FAU: 37, also I03_UBT: 37, 80). In some cases, the individual derivations proved to be problematic, as they did not always correctly reveal the actual meaning. This became particularly clear regarding the term “computational thinking”: “I would actually assume that it means artificial intelligence. (....) However, after reading through the answer options, I am no longer sure.” (I02_FAU: 8; also I04_UBT: 79 or I05_UBT: 52). Uncertainties arose from terms that were difficult for the students to understand, which made individual assignment to the competence areas more complicated: “Exactly, there were terms in it that I heard for the first time, (...) They were incomprehensible to me and so I didn't actually understand the meaning and content of the question itself and so I couldn't answer it.” (I01_UBT: 12). The problems described regarding the terminology used provide very specific indications of opportunities for further development.

Answer formulation

The feedback on the competence descriptions given (response items) was categorized according to positive and negative statements, as well as suggestions for improvement. Furthermore, “negative statements” were inductively divided into four additional sub-sub-categories: *difficulties in understanding*, *uncertainties*, *wording unsuitable for the target group*, *hierarchical structure of the competence descriptions and wording too long*. Thus, the sub-sub-category *difficulties in understanding* was coded most frequently. Seven student teachers commented on this with varying degrees. The use of technical terms, such as “digital artifacts” (I04_UBT: 123), caused difficulties. This made it difficult to understand and gave the impression that “it was about something incredibly complicated.” (I04_FAU: 30, 60, similarly I04_UBT: 147). Succinctly worded headings only appeared comprehensible after reading the answer guidelines (I05_UBT: 52, 74-75). The questions had to be read several times together with competence descriptions and examples to be able to answer them (I03_FAU: 30). In addition to specific formulations, the statements of five of the students highlighted a further difficulty: the hierarchically structured competence levels. In particular, the fact that

each higher competence level is supposed to include all lower ones poses challenges for the students. It was described as “stupid”, “problematic” or “difficult” (I01_UBT: 92, I01_UBT: 34, I02_UBT: 80-81). Some students made it clear that the answer requirements were perceived as inconsistent: “It's like a giant leap now: ‘tried to look for digital resources once.’ and ‘I use different digital tools.’” (I05_UBT: 54, similarly also I04_UBT: 121). However, there was often a lack of detailed explanations as to why an answer did not seem comprehensible (I05_UBT: 27-28, 50, 58-59).

Motivation in answering

The mere fact that the students voluntarily agreed to participate in the survey suggests that they had a certain basic motivation and interest. However, even the information on the estimated duration of the processing time had a deterrent effect. If it had not been a survey situation, participation would have been terminated right at the beginning (I02_FAU: 5, 21, 8, 37). After a short time, two students commented that the length of the answer options was exhausting (I02_UBT: 17; I04_UBT: 90, 141). Two other respondents adapted their reception of the questionnaire and skimmed over the answers of the higher competence levels (I03_UBT: 41, 63; I04_FAU: 12, 70, 17, 25, 32, 37). In addition to the length, the answer requirements had a demotivating effect. For example, one student said, “I have to say, I'm beginning to wonder whether I'm suitable for the survey (...)” (I04_FAU: 38). Overall, the impression was created that very few students would have completed the self-assessment without a mandatory framework.

Emotion in answering

Student statements in which they verbalized their emotions revealed the diversity of feelings. For example, at the beginning, the self-assessment and the comparison with the actual result were rated as very interesting (I02_FAU: 9) and one question was rated as exciting “because although it (.) is geared towards everyday professional life, (.) it can also be asked in the private sphere (.)” (I02_FAU: 7). However, the increasing cross-reading made the questionnaire boring (I01_FAU: 31) and monotonous (I01_FAU: 31) and led to a loss of concentration (I04_FAU: 46, 58). A little later, the student expressed that it was “annoying (...) that you can't (.) really take part in the survey, (.) because you only have (.) three possible answers.” (I04_FAU: 41 like I02_UBT: 38). Fears were expressed that the result would reveal an “extreme incompetence” (I01_FAU: 31). Accordingly, some of the final assessments were also critical. One student summarized: “In summary, I don't like it at all. Well, aehm, to be honest, I didn't enjoy it that much either.” (I02_UBT: 59). Others were able to find positive aspects in the self-assessment. The examples of the higher competence level were perceived as interesting and a good stimulus in terms of how much can be achieved (I04_FAU: 40, 44, 34, 54).

Formatting

The formatting of a questionnaire appears to be a marginal issue in the research process. In terms of student feedback, however, it can significantly increase motivation and user-friendliness.

For example, the formatting supported orientation in the questionnaire. The progress bar and the consistent structure contributed to this, examples were written in italics, and keywords were highlighted (I01_UBT: 38, I01_UBT: 38;

I01_FAU: 39, 49, 53, 57; I02_FAU: 27). Additionally, students favored a different color scheme. Another suggested a change of colors depending on the area of competence. (I01_FAU: 57). Moreover, the font size was also described negatively. Three students stated that the text could appear slightly larger (I02_UBT: 85; I03_UBT: 63; I04_FAU: 72, 77). The scroll-over effect for difficult terms also did not appear to be self-explanatory. In some cases, this was simply overlooked (I03_UBT: 99; I01_UBT: 40; I02_UBT: 86-93). Understandably, the lack of perceived support greatly impairs the comprehensibility of the items. The effort of the self-assessment was perceived very differently. Some students found the length acceptable (I04_UBT: 161, I02_FAU: 27; I03_UBT: 103; I03_FAU: 34) for others the questionnaire seemed "mega extensive" (I04_UBT: 145).

Feedback on the evaluation received

After completing the questionnaire, students were asked to self-assess their competence once again. After that, they received graphical feedback of their individual results. This was perceived as positive – even exciting – by many students (I01_UBT: 4, 28; I03_UBT: 105; I04_UBT: 171; I03_FAU: 20, 44; I04_FAU: 49; I05_UBT: 105). However, this does not apply to everyone. One student teacher, for example, rated the evaluation as "completely irrelevant for him as a prospective teacher, because it doesn't tell me anything, because I am simply (.) forced to land at a low level due to the given answer options" (I01_FAU: 51). The evaluation itself gives an overview of the individual results for each competence area in the form of a summary. On the one hand, this provides guidance, but on the other hand some students were confused by the graphical feedback and asked "What is this supposed to tell me?" (I01_FAU: 59). Others also asked how they could improve their competences after having received their results (I04_FAU: 79).

Conclusions and recommendations for further development based on the qualitative results

In summary to RQ 2, it is clear from the students' comments that, despite the scope of the survey instrument, they see added value in being able to assess their digital competences themselves. However, the questionnaire proved to be challenging for various reasons. Seven key aspects emerged from the qualitative survey using the thinking aloud method:

- The self-assessment tool poses challenges for students in many places, as it takes the perspective of in-service teachers.
- Explanations of technical terms should not only be supplemented, but also made more visible e.g. through more eye-catching formatting.
- The logic of the hierarchical level of competence should be reconsidered.
- The students initially showed willingness to participate in the survey, but factors like estimated duration and length of answers caused demotivation, leading some to question their suitability for such surveys, indicating that without a mandatory requirement, few would have completed the self-assessment.
- The statements from student teachers conveyed a range of emotions, with some finding aspects like self-assessment and comparison intriguing while others found the process tedious, monotonous, and anxiety-provoking, highlighting a variety of perspectives on the experience.

- The formatting is an important aid to orientation and speed of response.
- The evaluation of results can be enhanced by additional explanations and references to opportunities for personal development.

Use of the SELFIEforTEACHERS for self-assessment of the media-related digital competences of student teachers - Mixed Methods approach (RQ 3)

After the presentation of quantitative and qualitative results in the previous sections, a result-based integration of both strands is carried out in the following. Figure 1 and table 6 show a comparison of selected quantitative and qualitative results and are conceptualized as side-by-side displays (Kuckartz, 2017). Based on Johnson et al. (2019), the joint display was developed along the following four steps: 1) Listing: selective listing of the qualitative and quantitative data 2) Matching: merging tabular presentation of the qualitative and quantitative data 3) Checking: checking the comparison and selecting meaningful data material 4) Pillar building: conclusions on findings.

Figure 1

Relationship between the difficulties expressed in answering and the results at sub-category level

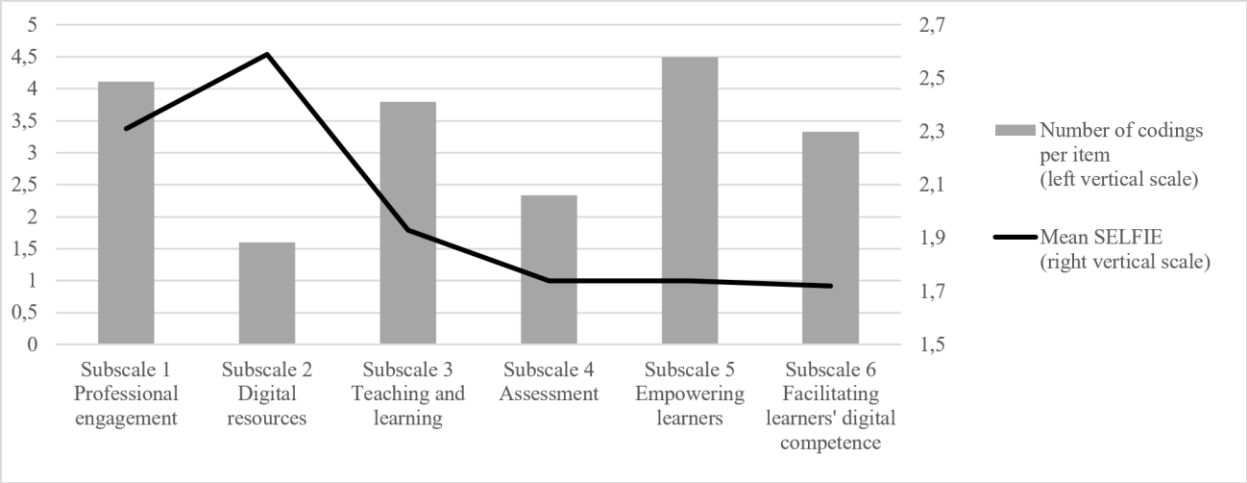

Figure 1 links the mean values of the results on the subscales (line chart) with the difficulties in answering (bar chart), which were recorded under subcode 2.2. A total of 109 codes were determined for processing difficulties that are directly related to SELFIEforTEACHERS. In the graph, the number of codes per subscale was standardised to the number of items in the respective subscale. For subscale 1, with 37 codes for 9 items, a high level of uncertainty can be reported. This can be explained by the novelty effect of the survey tool for the respondents. The result for the subscale is the second highest across all scales. There are significantly fewer difficulties with subscale 2. This is also due to the aforementioned composition of the thinking aloud group, in which every student has already attended at least one corresponding course. The average result of the SELFIEforTEACHERS in this sub-category is the maximum

across all categories. From sub-category 3 onwards, the difficulties in processing increase and the results in the sub-categories deteriorate. These difficulties can no longer be explained by the novelty effect of the survey tool. If one also considers that the processing time decreases continuously from sub-category 2 onwards (cf. section Results - Sample), this indicates that the formulations of the higher competence levels are not adequate for student teachers and that they therefore assess themselves at a lower level at an early stage.

Sub-category 4.2, which addresses the difficulties of assessing the level of competence and the inappropriateness of the hierarchical structure, shows a similar trend. The qualitative analysis reveals that Area 2 *Digital Resources* have the most codes, which reflects uncertainties and difficulties in assessing one's own level of competence. Quantitative results support this as subscale 2 has the highest average processing time per item (table 6).

Table 6

Side-by-side display (summary of results)

	Qualitative Category	**Example of interviews**	**Quantitative Results**
Student perspective	(2) Answerability from student perspective (sub-category 2.2)	*"In my opinion, this item addresses once again the everyday life of in-service teachers and can't be answered appropriately from student teachers. (...) In general, there was hardly any references for student teachers in area 3."* (I02_FAU: 8)	Decreasing results in mean values from subscale 2 onwards.
Cumulative competence levels	(4) Answer formulation (sub-category 4.2)	*"There is a giant leap [between] 'I have tried searching for digital resources' and 'I use various digital tools [to search for digital resources] (...) and adapt these directly to the students' (...) I think that's a big leap."* (I05_FAU: 54-56)	Highest processing time for subscale 2 (38.47 sec.)
Shrinking motivation	(5) Motivation in answering (sub-category 5.3)	*"At some point, I stopped reading all answer options because I thought to myself, it won't fit anyway."* (I03_UBT: 63)	Decreasing processing time from subscale 2 onwards.
Emotion towards self-assessment	(6) Emotion in answering (sub-category 6.2)	*"But I'm afraid that my result is already that I'm extremely incompetent."* (I01_UBT: 31)	Sign. difference in self-assessment before (t1) and after (t2) using SELFIEforTEACHERS (M_{t1} = 2,35, M_{t2} = 2,17)

Moreover, the fact that mean results for subscale 2 are positively correlated with the number of courses attended may deliver another explanation for difficulties in assigning to a competence level. Unlike in other areas of the instrument where students do not feel addressed and automatically allocate themselves to a lower competence level, they can refer to their previous seminar experience here leading to a deeper examination of a broader spectrum of competence levels and to further assignment problems.

The codings from sub-category 5.3 give proof of a decreasing motivation causing a the rather superficial answering of the questions. This superficiality goes hand-in-hand with a diminishing willingness among the participants to read all answer options (I03_UBT: 63). Additionally, from Area 3 *Teaching and Learning* onwards, the number of codes in the qualitative analysis increases indicating a decline in motivation and more superficial processing of the questions. From a quantitative point of view these findings can be supported as a decreasing average processing time per item from subscale 2 onwards is reported (table 6).

Finally, qualitative results indicate fears and concerns that students may have overestimated their abilities and are now facing disappointment (I01_FAU: 31). Subcode 6.2, which deals with students' negative emotions, shows a deterioration in the POST assessment compared to the PRE assessment. These results are supported by the quantitative analysis, which shows a poorer assessment of competences after completing the tool.

To sum up, it can be stated that the integration of the qualitative and quantitative strands led to a partial validation of results. This validation is visualized in the juxtaposition of results in form the side-by-side displays. In the following, these findings are discussed to provide a comprehensive answer to RQ 3.

Discussion

The analyzed quantitative and qualitative data from the SELFIEforTEACHERS revealed some difficulties regarding the self-assessment tool, which are closely related to the target group of student teachers. Thus, several aspects of the tool are not optimally tailored to this group and impair its theoretical and statistical validity as well as its practical applicability. The main problems are identified below and will be discussed against the background of RQ 3 followed by suggestions for possible revision.

Student teachers may find it difficult to understand some item formulations of the SELFIEforTEACHERS tool. Category 3 revealed that there were serious comprehension problems with terms such as *Computational Thinking* for instance. In addition, problems also occurred when the items referred to practical and pedagogical experiences at school, which many student teachers still lacked. Due to this rather one-sided emphasis on the school context, this may lead to a neglection of important aspects of student teachers' training aside from school. Thus, student teachers also need to develop competences in other areas, such as pedagogical theory and research. Moreover, the length of some items in SELFIEforTEACHERS may be overwhelming and affect the precision of the self-assessment. Long and complex items may require more time and cognitive resources from student teachers, which leads to a shrinking motivation causing superficial or inaccurate responses (see table 6). Another point of discussion is the fact that SELFIEforTEACHERS directly refers to DigCompEdu as the underlying theoretical model, causing problems when transferring the inherent progression model onto the instrument. As mentioned above, Redecker (2017) used the Common European Framework of Reference for Languages (CEFR) and Bloom’s revised taxonomy (Anderson & Krathwohl, 2001) as guidance for elaborating the six stages of progression. However, neither the taxonomy models of the CEFR nor of Anderson and Krathwohl (2001) are described in a way that would imply an equidistance between

the cognitive stages. Instead, Redecker (2017) acknowledges with reference to the CEFR that although "the levels A1 and A2, B1 and B2 and C1 and C2 are closely related, there is a cognitive leap between A2 and B1 and B2 and C1 respectively. This is also true for the DigCompEdu competence progression." (p. 28). From a statistical point of view, this would have required the use of ordinal scaled data instead of metric data in SELFIEforTEACHERS implying the use of median values instead of sum scores. Furthermore, the use of sum scores for the assessment of teachers' digital competences in SELFIEforTEACHERS leads to an overrepresentation of those competence areas (e.g. Area 1: *Professional Engagement*) with many items over those with less items (e.g. Area 4: *Assessment*). Along with the applied use of sum scores, the assumption of a reflective measurement model needs to be reconsidered. Reflective measurement models understand the measurement indicator items as consequence of the underlying construct (Weiber & Mühlhaus, 2014), whereby the indicators are supposed to correlate strongly with each other in terms of form and content. Although the latter can be confirmed empirically regarding the SELFIEforTEACHERS (see *Cronbach's Alpha* values above), the applied strategy of building sum scores does not fit with reflective measurement models. Instead, it is a common strategy used in the context of formative measurement models, where usually an index is created determining which indicator variables are to be weighted and summed up (Döring & Bortz, 2016). The logic underlying formative measurement models is that the construct to be measured is the effect or consequence of the characteristic indicators (ibid.). The indicators can be very different and do not necessarily have to correlate with each other.

With that being said, SELFIEforTEACHERS is concerned with serious validity issues not only in terms of theoretical aspects but also regarding statistics. Although, content validity – at least for in-service teachers – can be considered as satisfying as the item development was "based on […] desk research and consultations with experts and practitioners" (Economou 2023a, p. 19), validity in terms of the underlying measurement model, assumed datatyp, and progression model need to be questioned. Together with problematic item formulations and comprehension difficulties among the participants in this study it can be concluded that SELFIEforTEACHERS is only suitable for student teachers to a limited extent. To improve the effectiveness of the tool and extend its applicability to student teachers, urgent revision measures are required.

Limitations and Conclusion

Our empirical studies have shown that the SELFIEforTEACHERS self-assessment tool, which was originally developed for teachers, is not optimally tailored to student teachers. However, the results of our studies is limited due to the size of our sample. A larger and more diverse sample of student teachers could lead to more differentiated results. Furthermore, the generalization of our results might be limited by the specific institutional and cultural contexts of the teacher education programs studied. Further empirical studies in different educational institutions could provide a broader understanding. The student teachers' assessments while using SELFIEforTEACHERS may have been subjective and influenced by personal experiences and attitudes. This could also have led to distortions in the self-assessments.

Nevertheless, based on our findings, we propose a couple of revision approaches. First, it is necessary to adapt the item wording to the specific needs and experiences of student teachers. This could be done by revising the language and adapting the questions to ensure that they are clear and understandable for student teachers. In addition, the items should also cover aspects of education that go beyond the school context to capture a comprehensive picture of students' competences. In order to counteract the decreasing motivation and concentration of the participants, it would be conceivable to enable independent processing of individual sub-areas and to report interim results. Moreover, a renewed in-depth discussion and reconsideration about the underlying measurement model of the SELFIEforTEACHERS tool is necessary. Depending on the outcome of this debate, different ways of revision can be considered. In general, we propose the introduction of a five-point Likert scale after each item statement. In case of maintaining the reflective measurement model, a five-point Likert scale would deliver more precise results because participants would have to mark their individual level of agreement to each item statement instead of choosing one statement out of six as in the current version of the instrument. Rather than achieving a fixed numerical value depending on the chosen answer, mean values should be used to represent the extent of approval for each item and/or sub-category. From our viewpoint, this would allow student teachers to evaluate their competences in a more differentiated way without forcing them into a hierarchical structure. Assuming the existence of an underlying formative measurement model, a five-point Likert scale would work as well. Thus, participants would receive points for each statement (not points for selecting one statement out of six) according to their degree of approval. These points can be added up to a summative index score, if all items or subscales are considered as equally relevant for teachers' digital competences, or they can be calculated into a weighted index.

Overall, adapting the SELFIEforTEACHERS tool to the needs of student teachers requires a careful revision of both – item formulations and the underlying measurement model. By implementing the proposed revision approaches, the usefulness and relevance of the tool for this target group could be significantly improved. When revising the SELFIEforTEACHERS tool, it is important to consider different methodological approaches. This could include involving student teachers in the revision process as well as conducting further empirical studies to validate the proposed changes. A successful implementation of these proposed revisions may help to improve the quality and effectiveness of the SELFIEforTEACHERS to become a suitable research and self-assessment tool for student teachers.

References

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Bärnreuther, C. (2024). Awareness of Digital Competences among Student Teachers - A Mixed-Methods Study on the Role of Seminars Promoting Digital Competences. In T. Bastiaens (Ed.), *Proceedings of EdMedia + Innovate Learning* (pp. 440-449). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/224549/.

Blömeke, S. (2000). *Medienpädagogische Kompetenz: Theoretische und empirische Fundierung eines zentralen Elements der Lehrerausbildung*. kopaed.

Caena, F., & Redecker, C. (2019). Aligning teacher competence frameworks to 21st century challenges: The case for the European Digital Competence Framework for Educators (*Digcompedu*). *European Journal of Education*, *54*(3), 356–369. https://doi.org/10.1111/ejed.12345

Chai, C. S., Ng, E. M., Li, W., Hong, H.-Y., & Koh, J. H. L. (2013). Validating and modelling technological pedagogical content knowledge framework among Asian preservice teachers. *Australasian Journal of Educational Technology*, *29*(1). https://doi.org/10.14742/ajet.174

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th edition, international student edition). Sage.

DCB (Forschungsgruppe Lehrerbildung Digitaler Campus Bayern). (2017). Kernkompetenzen von Lehrkräften für das Unterrichten in einer digitalisierten Welt. *Merz Medien + Erziehung: Zeitschrift Für Medienpädagogik*, *4*, 65–74. https://www.researchgate.net/publication/319665661

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer. https://doi.org/10.1007/978-3-642-41089-5

Dresing, T., & Pehl, T. (Eds.) (2018). *Praxisbuch Interview, Transkription & Analyse: Anleitungen und Regelsysteme für qualitativ Forschende* (8th ed.). Dr. Dresing und Pehl GmbH.

Döbeli Honegger, B. (2021). Covid-19 und die digitale Transformation in der Schweizer Lehrerinnen- und Lehrerbildung. Beiträge zur Lehrerinnen- und Lehrerbildung, 39(3), 411–422.

Drossel, K., Eickelmann, B., Schaumburg, H. & Labusch, A. (2019). Nutzung digitaler Medien und Prädikatoren aus der Perspektive der Lehrerinnen und Lehrer im internationalen Vergleich. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil, & J. Vahrenhold (Eds.), *ICILS 2018 # Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (pp. 205–240). Waxmann.

Economou, A. (2023a). *SELFIEforTEACHERSs: Designing and developing a self-reflection tool for teachers' digital competence.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/561258

Economou, A. (2023b). *SELFIEforTEACHERS TOOLKIT: Supporting teachers in building their digital competence.* Publications Office of the European Union. https://data.europa.eu/doi/10.2760/626409

Endberg, M., & Lorenz, R. (2017.). Selbsteinschätzung medienbezogener Kompetenzen von Lehrpersonen der Sekundarstufe I im Bundesländervergleich und im Trend von 2016 bis 2017. In R. Lorenz, W. Bos, M. Endberg, S. Grafe, & J. Vahrenhold (Eds.), *Schule digital – der Länderindikator 2017: Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017* (pp. 151–177). Waxmann.

Endberg, M., & Lorenz, R. (2022). Selbsteingeschätzte Kompetenzen von Lehrpersonen der Sekundarstufe I zum Einsatz digitaler Medien im Unterricht im Bundesländervergleich 2021 und im Trend seit 2017. In R. Lorenz, S. Yotyodying, B. Eickelmann, & M. Endberg (Eds.), *Schule digital – der Länderindikator 2021: Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017* (pp. 11–21). Waxmann. https://doi.org/10.31244/9783830995449

Gysbers, A. (2008). *Lehrer - Medien - Kompetenz: Eine empirische Untersuchung zur medienpädagogischen Kompetenz und Performanz niedersächsischer Lehrkräfte*. Vistas-Verlag.

Ghomi, M., & Redecker, C. (2019). Digital Competence of Educators (DigCompEdu): Development and Evaluation of a Self-assessment Instrument for Teachers' Digital Competence. *Proceedings of the 11th International Conference on Computer Supported Education*, 541–548. https://doi.org/10.5220/0007679005410548

Guggemos, J., & Seufert, S. (2021). Teaching with and teaching about technology—Evidence for professional development of in-service teachers. *Computers in Human Behavior*, *115*, 106613. https://doi.org/10.1016/j.chb.2020.106613

Herzig, B., & Martin, A. (2018). Lehrerbildung in der digitalen Welt. In S. Ladel, J. Knopf, & A. Weinberger (Eds.), *Digitalisierung und Bildung* (pp. 89–113). Springer. https://doi.org/10.1007/978-3-658-18333-2_6

Herzig, B., Martin, A., Schaper, N., & Ossenschmidt, D. (2015). Modellierung und Messung medienpädagogischer Kompetenz—Grundlagen und erste Ergebnisse. In B. Koch-Priewe, A. Köker, J. Seifreid, & E. Wuttke (Eds.), *Kompetenzerwerb an Hochschulen: Modellierung und Messung: Zur Professionalisierung angehender Lehrerinnen und Lehrer sowie frühpädagogischer Fachkräfte* (pp. 153–176). Julius Klinkhardt.

Herzig, B., Schaper, N., Martin, A., & Ossenschmidt, D. (2016). *Schlussbericht zum BMBF Verbundprojekt M3K – Modellierung und Messung medienpädagogischer Kompetenz, Teilprojekt: Medienerzieherische und mediendidaktische Facetten und handlungsleitende Einstellungen.* Universität Paderborn.

Hofmann, M. (2017). *Lautes Denken. QUASUS. Qualitatives Methodenportal zur Qualitativen Sozial-, Unterrichts- und Schulforschung.* https://www.ph-freiburg.de/quasus/was-muss-ich-wissen/daten-erheben/lautes-denken.html

Huwer, J., Irion, T., Kuntze, S., Schaal, S., & Thyssen, C. (2019). Von TPaCK zu DPaCK – Digitalisierung im Unterricht erfordert mehr als technisches Wissen. *MNU Journal*, *70*(5), 358–364. https://www.researchgate.net/publication/335748844

Johnson, R. E., Grove, A. L., & Clarke, A. (2019). Pillar Integration Process: A Joint Display Technique to Integrate Data in Mixed Methods Research. *Journal of Mixed Methods Research*, *13*(3), 301–320. https://doi.org/10.1177/1558689817743108

Konrad, K. (2010). Lautes Denken. In G. Mey & K. Mruck (Eds.), *Handbuch Qualitative Forschung in der Psychologie* (pp. 476–490). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92052-8_34

Kuckartz, U. (2017). Datenanalyse in der Mixed-Methods-Forschung: Strategien der Integration von qualitativen und quantitativen Daten und Ergebnissen. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *69*(S2), 157–183. https://doi.org/10.1007/s11577-017-0456-z

Kuckartz, U. (2018). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung* (4th ed.). Beltz Juventa.

Kultusministerkonferenz (2021). *Lehren und Lernen in der digitalen Welt: Die ergänzende Empfehlung zur Strategie 'Bildung in der digitalen Welt'*. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2021/2021_12_09-Lehren-und-Lernen-Digi.pdf

Lorenz, R., Bos, W., Endberg, M., Eickelmann, B., Grafe, S., & Vahrenhold, J. (Eds.). (2017). *Schule digital – der Länderindikator 2017: Schulische Medienbildung in der Sekundarstufe I mit besonderem Fokus auf MINT-Fächer im Bundesländervergleich und Trends von 2015 bis 2017*. Waxmann.

Lorenz, R., Eickelmann, B., Endberg, M., & Yotyodying, S. (2022). Schule digital – der Länderindikator 2021: Theoretisches Rahmenmodell und Überblick über zentrale Ergebnisse. In R. Lorenz, S. Yotyodying, B. Eickelmann, & M. Endberg (Eds.), *Schule digital – der Länderindikator 2021: Lehren und Lernen mit digitalen Medien in der Sekundarstufe I in Deutschland im Bundesländervergleich und im Trend seit 2017* (pp. 89–115). Waxmann. https://doi.org/10.31244/9783830995449

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Quast, J., Rubach, C., & Lazarides, R. (2021). Lehrkräfteeinschätzungen zu Unterrichtsqualität mit digitalen Medien: Zusammenhänge zur wahrgenommenen technischen Schulausstattung, Medienunterstützung, digitalen Kompetenzselbsteinschätzungen und Wertüberzeugungen. *Zeitschrift für Bildungsforschung*, 11, 309–341. https://doi.org/10.1007/s35834-021-00313-7

Quast, J., Rubach, C., & Porsch, R. (2023). Professional digital competence beliefs of student teachers, pre-service teachers and teachers: Validating an instrument based on the DigCompEdu framework. *European Journal of Teacher Education*, 1–24. https://doi.org/10.1080/02619768.2023.2251663

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu.* Luxembourg: Publications Office of the European Union. https://doi.org/10.2760/159770

Rubach, C., & Lazarides, R. (2023). A Systematic Review of Research Examining Teachers' Competence-Related Beliefs About ICT Use: Frameworks and Related Measures. In K. Scheiter & I. Gogolin (Eds.), *Bildung für eine digitale Zukunft* (Vol. 15, pp. 189–230). Springer. https://doi.org/10.1007/978-3-658-37895-0_8

Sailer, M., Stadler, M., Schultz-Pernice, F., Franke, U., Schöffmann, C., Paniotova, V., Husagic, L., & Fischer, F. (2021). Technology-related teaching skills and attitudes: Validation of a scenario-based self-assessment instrument for teachers. *Computers in Human Behavior*, *115*, 106625. https://doi.org/10.1016/j.chb.2020.106625

Schmid, M., Brianza, E., Mok, S. Y., & Petko, D. (2024). Running in circles: A systematic review of reviews on technological pedagogical content knowledge (TPACK). *Computers & Education*, *214*, 105024. https://doi.org/10.1016/j.compedu.2024.105024

Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, *157*, 103967. https://doi.org/10.1016/j.compedu.2020.103967

Schmid, M., Brianza, E., & Petko, D. (2021). Self-reported technological pedagogical content knowledge (TPACK) of pre-service teachers in relation to digital technology use in lesson plans. *Computers in Human Behavior*, *115*, 106586. https://doi.org/10.1016/j.chb.2020.106586

Schmid, M., & Petko, D. (2020). ‹Technological Pedagogical Content Knowledge› als Leitmodell medienpädagogischer Kompetenz. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 121–140. https://doi.org/10.21240/mpaed/jb17/2020.04.28.X

Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological Pedagogical Content Knowledge (TPACK): The Development and Validation of an Assessment Instrument for Preservice Teachers. *Journal of Research on Technology in Education*, *42*(2), 123–149. https://doi.org/10.1080/15391523.2009.10782544

Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. *Zeitschrift für Erziehungswissenschaft*, *24*, 1039–1060. https://doi.org/10.1007/s11618-021-01047-y

Tiede, J. (2020). *Media-related Educational Competencies of German and US Preservice Teachers. A Comparative Analysis of Competency Models, Measurements and Practices of Advancement* [Dissertation, Julius-Maximilians-Universität Würzburg]. MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung. https://www.medienpaed.com/issue/view/84

Tiede, J., Grafe, S., & Hobbs, R. (2015). Pedagogical Media Competencies of Preservice Teachers in Germany and the United States: A Comparative Analysis of Theory and Practice. *Peabody Journal of Education*, *90*(4), 533–545. https://doi.org/10.1080/0161956X.2015.1068083

Tulodziecki, G. (2012). Medienpädagogische Kompetenz und Standards in der Lehrerbildung. In R. Schulz-Zander, B. Eickelmann, H. Moser, H. Niesyto, & P. Grell (Eds.), *Jahrbuch Medienpädagogik 9* (pp. 271–297). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-94219-3_13

Tulodziecki, G. (2021). Medienpädagogische Kompetenz. In U. Sander, F. Von Gross, & K.-U. Hugger (Eds.), *Handbuch Medienpädagogik* (pp. 1–9). Springer. https://doi.org/10.1007/978-3-658-25090-4_101-1

Valtonen, T., Sointu, E. T., Mäkitalo-Siegl, K., & Kukkonen, J. (2015). Developing a TPACK measurement instrument for 21st century pre-service teachersi. International Journal of Media, Technology and Lifelong Learning, 11(2).

Vejvoda, J., Stadler, M., Schultz-Pernice, F., Fischer, F., & Sailer, M. (2023). Getting ready for teaching with digital technologies: Scenario-based self-assessment in teacher education and professional development. *Unterrichtswissenschaft*, *51*(4), 511–532. https://doi.org/10.1007/s42010-023-00186-x

Vejvoda, J., Stadler, M., Schultz-Pernice, F., Glas, J., Fischer, F., & Sailer, M. (2024). Messung instrumenteller und kritisch-reflexiver Medienkompetenzen sowie medienbezogener Lehrkompetenzen von (angehenden) Lehrkräften: Entwicklung einer szenarienbasierten Kurzskala. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, *57*, 103–126. https://doi.org/10.21240/mpaed/57/2024.03.28.X

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Springer. https://doi.org/10.1007/978-3-642-35012-2

Yurdakul, I. K., Odabasi, H. F., Kilicer, K., Coklar, A. N., Birinci, G., & Kurt, A. A. (2012). The development, validity and reliability of TPACK-deep: A technological pedagogical content knowledge scale. *Computers & Education*, *58*(3), 964–977. https://doi.org/10.1016/j.compedu.2011.10.012

Corresponding Author Contact Information:

Author name: Christian Seyferth-Zapf

Department: Digital Teaching and Learning and Didactics of Computer Science

University, Country: Bayreuth, Germany

Email: christian.seyferth-zapf@uni-bayreuth.de

Please Cite: Seyferth-Zapf, C., Bärnreuther, C., Stephan, M., Ehmann, M., & Seyferth-Zapf, M. (2024). SELFIEfor(STUDENT)TEACHERS? A Mixed-Methods Study on the Suitabiliy and Validity of the Self-Reflection Tool for Student Teachers. *The European Educational Researcher*, *7*(3), 7-33. DOI: https://doi.org/10.31757/euer.732

Conflict of Interest: The authors declare no conflict of interest. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript**.**

Data Availability Statement: Data are available on request from the author.

Ethics Statement: Participants of this study gave their consent to use the data for research.

Author Contributions: Study conception, design, material preparation, and execution as well as data collection were done by all authors. Quantitative Data analysis was performed by Christian Seyferth-Zapf and Matthias Ehmann. Qualitative Data analysis was carried out by Cindy Bärnreuther, Melanie Stephan and Maria Seyferth-Zapf. All authors contributed to the manuscript to varying degree.

Received: March 8, 2024 ▪ Accepted: July 22, 2024