



Development and Application of a Domain-Specific TPACK Questionnaire—Findings from a Longitudinal Study on Teaching Human Biology Using Digital Tools

Leona Mölgen¹ · Roman Asshoff² · Benedikt Heuckmann²

Accepted: 8 March 2024
© The Author(s) 2024

Abstract

There are several specific digital tools now that have transformed the way science is taught. Correspondingly, teacher education programs have changed, and now they increasingly address the development of technology-related professional content knowledge (TPACK). Owing to the use of technology in specific domains of science teaching, there is an emergent need for domain-specific TPACK questionnaires. The present study investigates the development and application of a domain-specific TPACK questionnaire for teaching human biology using digital tools in a university-based teacher education program. A quasi-experimental between-subject design was applied for 13 weeks in the four study groups ($n = 155$). The intervention groups worked with digital tools in human biology, while the control groups participated without specific consideration of digital tools throughout the program. We succeeded in developing a questionnaire comprising seven reliable scales. The questionnaire development procedure described herein may be applicable to other science education disciplines. By the means of linear mixed modeling, we found that all students gained professional knowledge; however, treatment \times time interactions revealed that belonging to the intervention group was the main driver of TPACK progression. Comparing the treatments, we found that at post-test, the intervention group reported higher levels of TPCK, TPK, and TCK, but not CK, PK, TK, and PCK. Sequential analyses of the longitudinal data highlighted that working with digital tools on a regular basis constitutes a TPACK boost effect. This study contributes to the literature on technology integration from the perspective of biology education, where specific tools that can be assessed using a domain-specific questionnaire are used.

Keywords TPACK · Professional development · Technology integration · Human biology

Introduction

The constant development of educational technologies offers new approaches to accessing scientific phenomena using digital tools (Rosenberg & Koehler, 2015; Tan et al., 2022). In science teaching, digital technologies such as augmented and virtual reality (Cheng & Tsai, 2013; Lindgren et al., 2016), tools for digital data acquisition and processing, and 3D modeling (i.e., Thoms et al., 2022) are widely available and accessible. The potential of these technologies and

tools to transform science education is certainly not in doubt (European Commission, 2020). Consequently, integrating them into science teaching sets new demands for science teachers' professional knowledge and skills (von Kotzebue, 2022a, b).

In science teacher education, theoretical frameworks on teachers' professional knowledge (Blömeke et al., 2015; Hume et al., 2019; Shulman, 1986) have been modified to include technology-related aspects. For example, the TPACK¹ framework by Mishra and Koehler (2006) expanded Shulman's (1986) PCK model by a knowledge domain that explicitly addresses technological knowledge. Such frameworks provide a theoretical basis for the

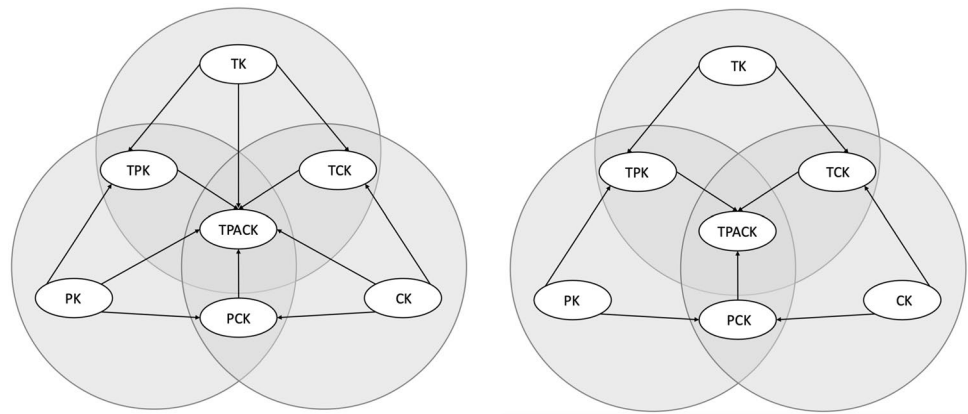
✉ Benedikt Heuckmann
Benedikt.Heuckmann@uni-muenster.de

¹ Institute for Science Education, Leibniz University Hannover, Am Kleinen Felde 30, 30167 Hannover, Germany

² Centre for Biology Education, University of Münster, Schlossplatz 34, 48143 Münster, Germany

¹ Please note that we are using the acronym TPACK to refer to the complete framework of technology-related pedagogical content knowledge as conceptualized by Koehler and Mishra (2009) and the acronym TPCK to refer to the singular construct or scale.

Fig. 1 Integrative and transformative view of TPACK components (figure adapted; based on an idea by Schmid et al., 2020)



development of questionnaires designed to measure teachers' knowledge and skills at a general level. As a prominent example, Schmidt et al. (2009) developed a TPACK questionnaire that has been applied in numerous studies (Chai et al., 2013). It is primarily of use for studies investigating teachers' TPACK and TPACK-related outcomes in teacher education (i.e., cross-sectional studies). Less often, TPACK questionnaires are used to investigate teachers' professional development (i.e., longitudinal studies; Wang et al., 2018). However, this would be important to better understand how teacher education programs contribute to TPACK development and how teachers can be better prepared to integrate technology into their teaching.

A prerequisite for the accurate measurement of TPACK professional development is the availability of questionnaires that allow for reliable and valid inferences to be drawn in the specific domain under investigation (Taber, 2018). Compared to research on PCK, where the development of domain-specific questionnaires has a long history (i.e., Großschedl et al., 2019), there are relatively few instruments available for assessing domain-specific TPACK (Tan et al., 2022). However, given the trend of technology becoming more prevalent in teacher education, the accurate measurement of TPACK and its development becomes an ever more important part of science education research (von Kotzebue, 2022b).

In the present study, we intend to combine the perspectives of developing a domain-specific TPACK questionnaire and survey longitudinal TPACK development throughout a university-based teacher education program. The contextual background of this study is biology education, and the domain under investigation is human biology. We focused on this domain because it is obligatory for science teaching and (biology) teacher education in most countries. Moreover, there is a wide availability of digital technology and tools that transform the way human biology is taught in schools (Chakraborty & Cooperstein, 2018; Lewis et al., 2014; Moro et al., 2017). The study is of interest to science education

researchers interested in TPACK research, the development of domain-specific TPACK questionnaires, and the longitudinal development of TPACK.

Theoretical Background

The TPACK Framework and Teachers' Professional Development

The TPACK framework has repeatedly been used to conceptualize the professional knowledge of teachers who research science teacher education and technology (Angeli & Valanides, 2015; Herring et al., 2016; Tan et al., 2022). The consideration of technology-related knowledge added new types of knowledge to the discourse on PCK (Rosenberg & Koehler, 2015; Schmid et al., 2020). In addition to pedagogical knowledge (PK), content knowledge (CK), and pedagogical content knowledge (PCK) (Hume et al., 2019), Mishra and Koehler's (2006) TPACK framework complements four technology-related types of knowledge: technological knowledge (TK), technology-related pedagogical knowledge (TPK), technology-related content knowledge (TCK), and technology-related pedagogical content knowledge (TPCK). Following an integrative view (see the left side of Fig. 1), the different knowledge types (PK, CK, TK, TPK, TCK, and PCK) are combined to TPCK (Schmid et al., 2020). That is, high levels of knowledge in PK, CK, TK, TPK, TCK, and PCK should correspond to high levels of TPCK (von Kotzebue, 2022b). In contrast, following a transformative view of TPACK (see the right side of Fig. 1), knowledge at the intersections of PK, CK, and TK is transformed to new knowledge constructs—namely, TPK, TCK, and PCK—which consecutively form TPCK as a unique type of knowledge (Schmid et al., 2020; von Kotzebue, 2022b). Still, there is an ongoing debate about the factorial structure of TPCK. Empirical support exists for integrative and transformative views (Scherer et al., 2017; von Kotzebue, 2022a).

In response to the technological transformation of science teaching, professional development of teachers' TPCK

constitutes an emerging theme in science education research (Chai, 2019; Huang et al., 2022). In their systematic review of teachers' professional development in STEM, Huang et al. (2022) illustrated that teachers' technology-related (TK, TPK, TCK, TPCK) and non-technology-related knowledge (PK, CK, PCK) are commonly addressed during professional development. Combinations of TCK and TPCK as well as PCK and CK were the most frequent outcome variables. Moreover, Tan et al. (2022) emphasized that studies on the use of specific technology should be considered for TPACK development since specific technology is often applied in specific subject-matter contexts. Typically, specific courses or short online programs have been designed to foster teachers' knowledge gains (i.e., Doering et al., 2009; Niess et al., 2014; Rienties et al., 2013). However, studies that analyze TPACK development during regular domain-specific teacher education programs, e.g., at teacher college or university, are scarce (Lachner et al., 2021; Thoms et al., 2022).

Measuring Teachers' Professional Knowledge Regarding Technology Use

Several attempts have been made to assess teachers' professional knowledge and teachers' competencies in empirical science education research (Angeli & Valanides, 2009; Herring et al., 2016), with seminal frameworks that outline the competencies that teachers need to integrate digital technologies being available through the TPACK framework (Mishra & Koehler, 2006) and the UNESCO ICT competency framework for teachers (UNESCO, 2018).

Review studies (i.e., Abbitt, 2014; Chai et al., 2013; Rosenberg & Koehler, 2015; Wang et al., 2018) showed that most attempts to measure teachers' knowledge regarding technology use and integration follow the framework by Mishra and Koehler (2006). One of the most commonly used instruments measuring TPACK in this regard stems from Schmidt et al. (2009). The questionnaire was developed in a survey with primary school teachers during a 15-week course that addressed how to apply technology in mathematics, social studies, science, and literacy content areas in PK-6 classrooms. The instrument encompasses seven scales for the seven TPACK constructs with robust psychometric properties (i.e., 47 items in total, internal consistency range, $\alpha = .75-.92$). Similar to Schmidt et al. (2009), many fellow researchers have implemented Likert-type items that assess TPCK through self-reported data. This methodological approach is often criticized for solely measuring subjectively perceived knowledge and not representing a test-based approach to TPCK (i.e., Krauskopf & Forssell, 2018). However, a recent study by Schmid et al. (2021) on the relationship between self-reported TPACK and using technology in class revealed that self-reported data have some predictive validity for actual behavior, thus justifying its use. Since the original

questionnaire emerged in 2009, adapted versions are available, which reduce the number of items and administration time while maintaining psychometric quality (Angeli & Valanides, 2015; Schmid et al., 2020).

Domain-specific adaptations of the TPACK questionnaire to general contexts indicate that TPACK surveys can be successfully implemented in subject-specific contexts (Rosenberg & Koehler, 2015; Wang et al., 2018). In a study with pre-service teachers from five different subjects (i.e., mathematics and biology), Lachner et al. (2021) showed that TPACK questionnaires can be built using a combination of items that assess generic technological pedagogical knowledge (TPK) and items that assess subject-specific TPCK aspects. Furthermore, a group of science education researchers have developed a self-assessment tool based on the TPACK framework (DiKoLAN-Grid; von Kotzebue et al., 2021). It aims to measure subject-specific aspects of TPACK, namely, knowledge about special tools, content-specific contexts, methods and digitality, and teaching. For biology education, Mahler and Arnold (2022) developed a questionnaire that comprises all seven TPACK constructs. In their study of 403 student teachers, the authors showed the usability of the TPACK framework for German-speaking countries while taking biology-specific requirements into account.

Despite the promising developments of subject-specific TPACK questionnaires, there is still a research gap on TPACK questionnaires facilitated for use in domain-specific areas within subject contexts (Rosenberg & Koehler, 2015; Wang et al., 2018). What is missing is adaptation beyond the subject level to the level of specific domains, considering the role of contextual knowledge (Mishra, 2019). In a systematic review, Brianza et al. (2022) demonstrated that TPACK development is embedded into a specific context (micro-, meso-, and macrolevel) and a social, resource-oriented, or content-related dimension. The authors argue that taking the contextual nature into account can shed light on factors that influence TPACK development. In a similar vein, it is well known from research on teachers' PCK that domain-specific questionnaires are necessary to capture the nature of specific contextual knowledge (i.e., Großschedl et al., 2018, 2019). In terms of TPACK, specific domains require teachers to acquire specific knowledge and skills related to the educational technology used in this area (i.e., anatomy apps in the domain of human biology; Chakraborty & Cooperstein, 2018; Lewis et al., 2014; Moro et al., 2017; thermal cameras in the context of chemistry and physics education; Haglund et al., 2022).

Research Questions

In the present study, we focused on the development of a TPACK questionnaire for a specific domain of biology education, namely, human biology. To ensure broad

Table 1 Sample characteristics

#	Location	<i>n</i>	PoDC	Study group	Mean age (SD)	Teacher education program	Mean semester in study program (SD)	Credit points awarded	Students privately owning a tablet
Group 1	A	25	3	IG	23.13 (3.76)	Bachelor	4.28 (1.40)	3 ECTS	72.0%
Group 2	A	23	3	CG	22.09 (3.48)	Bachelor	4.48 (1.50)	3 ECTS	82.6%
Group 3	B	57	2	CG	21.39 (1.77)	Bachelor	3.21 (1.31)	1 ECTS	66.6%
Group 4	C	50	2	IG	24.11 (2.71)	Master	2.70 (2.90)	2 ECTS	58.3%
Total	N.A.	155	N.A.	N.A.	22.77 (3.50)	N.A.	N.A.	N.A.	67.3%

1 ECTS European Credit Transfer System, *1 ECTS* 30 h workload, *IG* intervention group, *CG* control group, *PoDC* points of data collection, *N.A.* not applicable

compatibility of the intended questionnaire with international research on TPACK, we aim to include all seven constructs of the TPACK framework and not just the constructs related to technology use. Thus, the first research question is as follows.

RQ1: To what extent does the questionnaire on TPACK related to teaching human biology with digital tools allow for the formation of reliable scales and draw valid inferences related to convergent and discriminant validity?

Furthermore, the present study takes TPACK professional development into consideration. This study stands in to fill the gap that most studies on TPACK professional development have been conducted in settings designed specifically for this purpose, (i.e., TPACK online programs, Hofer & Grandgenett, 2014; Wang et al., 2018) and examines how domain-specific TPACK develops in an authentic teacher education situation, that is, in the context of a regular teacher education program course. Thus, we formulated the second research question as follows.

RQ2: How does the TPACK develop during a period of 13 weeks (one semester) of a regular teacher education program on teaching human biology?

Methods

Procedures and Sample Description

This study applied a quasi-experimental between-subject design for a period of 13 weeks in four different study groups (see Table 1), two of them were intervention groups (group 1, group 4) and two were control groups (group 2, group 3). We did not pool the data of the intervention and control groups and compared the findings separately. Data were collected during a regular biology teacher education program dedicated to human biology using an online questionnaire so that the structural aspects of the contextual nature of

TPACK development are located at the microlevel and refer to the social dimension (i.e., pre-service teachers as actors) according to Brianza et al. (2022). In total, *n* = 155 biology teacher students from three different German universities voluntarily participated in the study. Table 1 presents the sample characteristics. Answering the questionnaire lasted approx. 10–15 min and took place between October 2021 and August 2022.

According to the national guidelines for university-based teacher education in Germany (KMK, 2019), human biology is a compulsory study content. Typically, student teachers attend lectures on human anatomy and physiology, followed by specific teacher-related educational courses that address how to teach the topic to school students. These courses focus on hands-on inquiry activities such as dissection of animal specimens (i.e., mammal eye), using anatomical models (i.e., skeleton), and conducting experiments related to human biology (i.e., protein digestion in the stomach). Based on science education and pedagogical theory, students discussed how these inquiry activities foster scientific literacy. In the present study, the same content was taught to all study groups. A detailed overview of the study content at the different locations is available as supplementary material (see Table S1, supplements). However, due to the different federal guidelines for teacher education in Germany, some groups allowed more time so that the content could be discussed in more detail (i.e., depending on the credit points awarded, see Table 1).

At the beginning of each course week, the students held a short stimulus presentation summarizing concepts relevant to the course theme (i.e., anatomy and physiology of the cardiovascular system). In the intervention groups (i.e., group 1 and group 4 in Table 1), students used tablet-based apps showing dynamic 3D models of human anatomy, physiology, and pathology for their presentation (Chakraborty & Cooperstein, 2018). These types of apps allow students to dynamically present anatomical and physiological content, for example, by zooming in and out to virtual bodies, adding and removing layers and body parts such as muscular or vascular tissue, skin,

and bones through live dissection activities, highlighting and annotating these models with 360° rotation, and using animation and simulation to show physiological processes (Lewis et al., 2014). In addition, the students could use the apps during all inquiry activities in class. For example, when dissecting animal specimens, students were encouraged to use digital 3D models in addition to the original specimen to explore the anatomy of the specimen. In contrast, in the control groups (i.e., group 2 and group 3 in Table 1), the students did not work with the apps. They used traditional digital presentation software (i.e., PowerPoint slides showing 2D pictures of anatomical models, mostly copied from books or internet resources) and had the opportunity to use analogue models of anatomical structures during all inquiry activities in class. In comparison, the digital tools used in the intervention groups go beyond solely substituting the traditional digital presentation software used in the control groups. Moreover, the apps help to augment, modify, or redefine working with anatomical models in science teacher education as described by the SAMR model (Hamilton et al., 2016; Lewis et al., 2014).

In all four groups, the lecturers held the presentations in the first few weeks of the semester (up to week 4 in groups 1 and 2 and up to week 3 in groups 3 and 4). This procedure was implemented so that the students have time become familiar with their topic of their presentation, prepare their presentations accordingly, and see good-practice examples how the lecturers want them to implement presentation in class.

The rationale behind the composition of study groups is threefold. First, we chose groups from three institutions in three federal states of Germany (see Table 1) to evaluate a diverse range of students and different federal curricula leading to the same degree (Bachelor of Education or Master of Education). The instructors responsible for delivering the courses agreed to follow the same curriculum on human biology to ensure comparability (Supplementary Table S1). Second, we selected groups enrolled in bachelor's and master's programs to investigate the teaching of human biology in both study programs, which is typical for teacher education in Germany. We hypothesized that master's students would have higher CK and PCK levels (Kleickmann et al., 2013). Additionally, we included the intervention and control groups at each study level. However, due to COVID-19 restrictions, we could not include a second group at the master's level so that the study group design has a disbalance here. Finally, we added an extra data collection point at week five in groups 1 and 2 (see Table 1) to examine TPACK development in more detail. Due to limitations in data assessment at locations B and C, we could only collect data three times at location A.

Part A: Development of the Questionnaire

We sourced items from the short TPACK assessment instrument (TPACK.xs) by Schmid et al. (2020) and adapted them to the specific domain of human biology for three main reasons: First, the instrument by Schmid et al. (2020) relies on to the original TPACK framework by Mishra and Koehler (2006) and encompasses all seven components of the TPACK framework. Hence, the instrument is eligible for the analysis of the transformative and integrative TPACK view (Schmid et al., 2020; von Kotzebue, 2022b). Second, TPACK.xs reduced the number of items from the initial instrument (Schmidt et al., 2009) from 42 to 28, thus reducing the time needed for data collection. A short administration time allowed us to add an intermediate point for data collection (see Table 1). Third, TPACK.xs has sound psychometric properties, allowing for a reliable and valid measurement of TPACK components. The instrument and its predecessor have proven its suitability for various domains and have successfully been applied to professional development studies (Huang et al., 2022; Schmid et al., 2021; Wang et al., 2018). For item adaptation, we systematically exchanged phrases referring to general contexts (i.e., “in my field,” “in my subject”) from the initial items by Schmidt et al. (2009) with phrases explicitly referring to the domain of human biology. This applies to all TPACK constructs that require contextual knowledge (i.e., CK, TCK, PCK, and TPCK). In addition, we changed phrases broadly referring to “technology” to “digital tools,” as this suits the intended use of applying tablet-based anatomy apps better. The term “digital tools” in contrast to digital technologies or information and communications technology (ICT) is used to refer to applying digital technologies and ICT for teaching and learning science, for example, for problem-solving, data processing, and enabling communication (Hillmayr et al., 2020).

Table 2 presents an overview of the scales and sample items used in the present study and contrasts them with the original items from the TPACK.xs instrument developed by Schmid et al. (2020). All items were measured on a five-point rating scale (1–5), with higher numbers representing stronger agreement with the statement provided (Robitzsch, 2020; Schmidt et al., 2009).

Part B: Data Analysis and Assessment Criteria

Regarding RQ1, we aggregated the items into scales according to the seven TPACK constructs described by Schmid et al. (2020). We calculated Cronbach's alpha (α) and McDonald's omega (ω) as measures of internal consistency and checked the item discriminatory power (r_{it}). As recommended by Zinbarg et al. (2005), we aimed for values of α and $\omega \geq .70$ and $r_{it} \geq .40$ to ensure the reliability of the scales. To assess the factorial validity of the scales,

Table 2 Overview of the TPACK scales and sample items used in the study and comparison with the original items from the TPACK.xs instrument

Scale	Sample item	Items
TK	I keep myself up to date on new developments of digital tools. (TK1) <i>I keep up with important new technologies. (TPACK.xs-TK1)</i>	3
CK	I have a sound knowledge in the domain of human biology. (CK1) <i>I have sufficient knowledge about my teaching subject. (TPACK.xs-CK1)</i>	4
PK	I can adapt my teaching style to the individual needs of the learners. (PK2) <i>I can adapt my teaching style to different learners. (TPACK.xs-PK1)</i>	4
TPK	I can critically reflect on using digital tools during my biology classes. (TPK4) <i>I am thinking critically about how to use technology in my classroom (TPACK.xs-TPK4)</i>	4
TCK	I am familiar with digital tools that can be used in the domain of human biology. (TCK3) <i>I know which new technologies are currently being developed in the field of my subject. (TPACK.xs-TCK3)</i>	4
PCK	I know how to promote learning in the domain of human biology by selecting effective teaching and learning strategies. (PCK1) <i>I know how to select effective teaching approaches to guide student thinking and learning in my teaching subject. (TPACK.xs-PCK1)</i>	4
TPCK	I can apply strategies to combine learning about human biology, digital tools, and pedagogical concepts in my biology classes in a meaningful way. (TPCK1) <i>I can use strategies that combine content, technologies, and teaching approaches that I learned about in my coursework in my classroom. (TPACK.xs-TPCK1)</i>	4

Original items by Schmidt et al. (2009) are in italics

TK technological knowledge, CK content knowledge, PK pedagogical knowledge, PCK pedagogical content knowledge, TCK technological content knowledge, TPK technological pedagogical knowledge, TPCK technological pedagogical content knowledge

we performed a series of confirmatory factor analyses (CFA, Kline, 2016) and calculated the factor loadings (λ) and model fit indices (chi-square values (χ^2), root square mean error of approximation (RMSEA), standardized root mean square residual (SRMR), Bentler comparative fit index (CFI), and Tucker-Lewis-index (TLI)). For model evaluation, values of $\chi^2/df \leq 3$, RMSEA and SRMR $\leq .08$ and CFI and TLI $\geq .90$, and $\lambda \geq .40$ were deemed acceptable based on Schreiber et al. (2006). First, we analyzed a 7-factorial model with each of the seven TPACK components being independent of each other. A 7-factorial model has been described by Mahler and Arnold (2022) as the TPACK model that fit best compared to models that differentiate between technology-related and non-technology-related aspects of TPACK. In the second step, we fit models specified as transformative and integrative views of TPCK (Schmid et al., 2020; von Kotzebue, 2022b). Third, we applied a chi-square-based model comparison test to find out which of the three models (independent 7-factorial, integrative, transformative) fit best to the data. For model evaluation, Akaike information criteria (AIC) and Bayesian information criteria (BIC) were computed. Lower information criteria indicate better-fitting models (Hoyle, 2012). We report the latent factor correlations of the best-fitting model to investigate to what extent convergent or discriminant valid inferences can be drawn from the data (Hoyle, 2012). According to Cohen (1992), $r = 1.10$ is represented as small, $r = 1.30$ as medium, and $r = 1.50$ as large correlation coefficient.

Regarding RQ2, we applied a series of linear mixed models (LMM) and specified treatment (intervention vs. control), time (pre-test, intermediate test, post-test), and group (study group 1 to 4) as fixed effects. We chose LMM over repeated measures analysis of variance since the latter cannot handle missing data and LMM takes care of nested data (i.e., group is nested in treatment and cohorts, see Table 1). LMM further allowed adding a random intercept (id) as random effect and this controls for individual differences regarding TPACK (Kuznetsova et al., 2017). We investigated the fitted LMM for main and interaction effects and calculated partial squared eta as effect size. We evaluated them according to the benchmarks of $\eta_p^2 = .01$ representing a small, $\eta_p^2 = .06$ a medium, and $\eta_p^2 = .14$ a large effect (Cohen, 1992). For post hoc tests, we calculated pairwise comparisons based on the estimated marginal means and used Tukey method for adjustment. For data analysis, we mainly used R Studio (R Core Team, 2021) as well as different R packages such as lavaan (Rosseel, 2012) and LME4 (Kuznetsova et al., 2017).

Results

Part A: Instrument Development

Overall, we succeeded in forming reliable scales as indicated by the measures of Cronbach's α and McDonald's ω (see Table 3). Each scale had high item discriminatory power (r_{it}) showing that the items aggregated in a scale strongly

Table 3 Reliability measures for all scales at pre-test, intermediate test, and post-test

Scales	α_{pre}	r_{it-pre}	ω_{pre}	α_{interm}	$r_{it-interm}$	ω_{interm}	α_{post}	r_{itpost}	ω_{post}
TK	.80	.60–.82	.80	.78	.53–.81	.80	.69	.53–.77	.75
PK	.80	.58–.86	.83	.78	.60–.80	.81	.71	.42–.84	.79
CK	.74	.55–.75	.77	.76	.54–.76	.79	.68	.37–.75	.74
TPK	.90	.67–.90	.93	.87	.68–.85	.91	.87	.67–.87	.90
TCK	.85	.72–.81	.90	.85	.72–.85	.89	.83	.66–.78	.88
PCK	.88	.72–.86	.88	.88	.75–.80	.91	.83	.66–.79	.87
TPCK	.91	.79–.89	.95	.87	.75–.82	.91	.90	.79–.84	.93

α Cronbach’s alpha, ω McDonald’s omega, r_{it} item discriminatory index, *pre* pre-test, *interm.* intermediate test, *post* post-test

correlate with each other. There was a decrease of the measures for α from pre-test to post-test in TK, PK, and CK indicating that the participants in the post-test answered these items in a different manner. However, we still deemed these scales to be sufficiently reliable since the corresponding measures of ω were well above the suggested benchmark.

Regarding factorial validity, the independent model ($\chi^2/df = 1.39$; CFI = .95; TLI = .94; RMSEA = .05, 90% CI [.04–.07]; SRMR = .06), the integrative model ($\chi^2/df = 1.42$; CFI = .95; TLI = .93; RMSEA = .06, 90% CI [.04–.07]; SRMR = .08), and the transformative model ($\chi^2/df = 1.42$; CFI = .95; TLI = .94; RMSEA = .06, 90% CI [.04–.07], SRMR = .08) met the benchmarks of acceptable model fit. All items significantly predicted the latent variables of the corresponding scales, and the factor loadings were between $\lambda = .56$ and .93. Thus, factorial validity can be established.

Regarding model comparison, the independent model fit significantly better than the integrative model ($\Delta_{chisq.} = 16.74$; $\Delta_{Df} = 6$; $p = .01$) and the transformative model ($\Delta_{chisq.} = 19.37$; $\Delta_{Df} = 9$; $p = .02$). Data from the present study favored neither the integrative nor the transformative model ($\Delta_{chisq.} = 2.66$; $\Delta_{Df} = 3$; $p = .44$).

Findings from the latent correlation analysis (Table 4) showed that the TPACK constructs are systematically related to each other. For example, the correlation coefficients between TK, CK, and PK scales were considerably low ($r = .29$ to $.32$). In a similar vein, scales that share two constructs but were expected to measure different constructs only correlated at low to medium magnitude (e.g., TK and

TPCK, $r = .46$). We considered these findings a suggestion that discriminant valid inferences can be drawn from the data. In contrast, the correlation coefficients between scales that have conceptual similarities (i.e., share two knowledge aspects) were elevated and of moderate magnitude (e.g., TCK and TK, $r = .55$; CK and PCK, $r = .53$; TPK and PK, $r = .64$). Other correlations (e.g., PCK and PK, $r = .79$; TPCK and PCK, $r = .79$) were of high magnitude, indicating that these constructs share substantial variance but still have unique aspects. We considered these findings a sign of convergent validity.

Part B: Longitudinal Development of TPACK

Figure 2 provides an overview of the TPACK development for all groups over the 13 weeks of intervention. The figure highlights variability in the students’ responses in terms of baseline responses (pre-test) and final responses (post-test). As a coherent pattern, pairs of study groups 1 (IG) and 2 (CG) as well as study groups 3 (CG) and 4 (IG) show a rather similar trend. However, the plots suggest that students in the IG have a stronger TPACK progression over time than students in the CG. Table 5 summarizes the findings on the corresponding main and interaction effects for the factors time, treatment, and group from LMM as well as the corresponding effect sizes.

For TK, there was a significant main effect for time, which revealed an increase of TK for all study groups

Table 4 Correlations between the latent scales

Scales	TK	CK	PK	PCK	TCK	TPK	TPCK
TK	–	.31	.29	.33	.55	.68	.46
CK		–	.56	.53	.41	.51	.59
PK			–	.79	.41	.64	.66
PCK				–	.50	.63	.79
TCK					–	.64	.63
TPK						–	.74
TPCK							–

All correlations were significant at $p < .001$ level

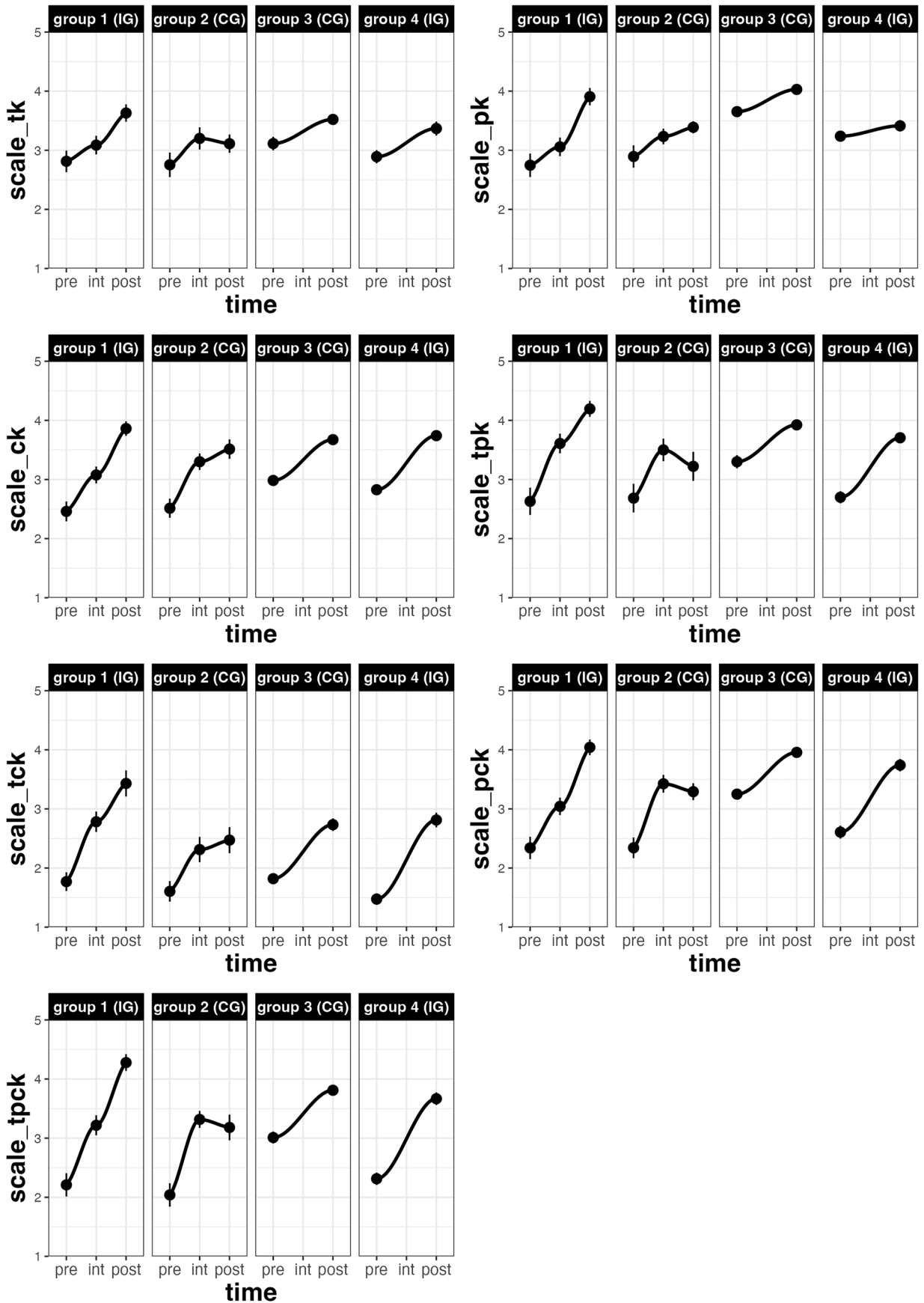


Fig. 2 Development of TPACK constructs over time in the four study groups

over time. For TCK, there was a significant main effect for group and an interaction effect for time × treatment. The main effect highlights that the groups on average over the levels of time expressed different TCK, with students in group 1 having a higher TCK compared to the other groups. Post hoc analyses for the interaction effect showed that the IG and CG expressed similar TCK at pre-test, $t(253) = -.47, p = .64$, but students in the IG reported higher TCK at post-test, $t(275) = 3.37, p < .001$.

For all other scales, we found significant interaction effects for time × treatment and for time × group indicating that the students’ expression of CK, PK, TPK, PCK, and TPCK varied over time depending on the type of treatment and study group (see Table 2).

For CK, post hoc analyses showed that while CK was not different between treatments at pre-test, $t(241) = -.89, p = .37$, students in the IG reported significant higher CK than students in CG at post-test, $t(268) = 1.98, p < .05$. Furthermore, post hoc analyses showed that all study groups significantly increased their CK over time ($p < .001$ for all comparisons pre–int, int–post, pre–post) except for group 2, which showed no significant progression between intermediate test and post-test, $t(152) = -1.68, p = .88$.

For PK, post hoc analyses showed that while PK was significantly higher in the CG than in the IG at pre-test, $t(247) = -2.45, p < .05$, there were no significant differences at post-test $t(271) = -.43, p = .67$. Furthermore, a pre–post comparison yielded significant PK progression for study group 1, $t(149) = -7.89, p < .001$, and for study group 3, $t(165) = -3.48, p < .05$, but not for study groups 2 and 4.

For TPK, post hoc analyses showed that while TPK was significantly higher in the CG compared to the IG at pre-test, $t(262) = -2.35, p < .05$, TPK turned out to be significantly lower in the CG at post-test, $t(279) = 2.2, p < .05$. Furthermore, there was a significant increase of TPK from pre-test to post-test for study groups 1, 3, and 4, but not for study group 2. In addition, analyzing the intermediate–post-test development revealed no significant changes of TPK in study group 1, $t(153) = -3.07, p = .10$, as well as in study group 2, $t(157) = .91, p = .99$.

For PCK, the data revealed an increase in PCK for both treatments from pre-test to post test, but the post-test differences are not statistically significant, $t(274) = 1.84, p = .07$. Furthermore, post hoc analyses showed that all study groups significantly increased PCK over time ($p < .001$ for all comparisons pre–int, int–post, pre–post) except for group 2, which showed no significant progression between intermediate test and post-test, $t(153) = .01, p = 1.00$, indicating a medium effect for time × group, $\eta_p^2 = .09$.

For TPCK, students in both treatment groups reported similar levels of TPCK at pre-test, $t(258) = 1.91, p = .06$, but students in the IG expressed significantly higher TPCK at post-test, $t(377) = 3.13, p < .01$. Furthermore, post hoc analyses showed that all study groups significantly increased TPCK over time ($p < .001$ for all comparisons pre–int, int–post, pre–post) except for group 2, which showed no significant progression between intermediate test and post-test, $t(157) = .31, p = 1.00$.

Summary of the Results

For all TPACK scales, students reported an increase in their agreement with the statements provided when comparing pre-test and post-test findings over the 13 weeks

Table 5 Summary of the main and interaction effects from the LMM

DV in LMM	Main and interaction effects	F-statistic	Effect size η_p^2
TK	Time	$F(153) = 23.96, p < .001$.20
PK	Treatment × time	$F(150) = 11.26, p < .001$.13
	Treatment × group	$F(160) = 14.81, p < .001$.16
CK	Treatment × time	$F(154) = 15.02, p < .001$.16
	Treatment × group	$F(164) = 8.38, p < .001$.09
TPK	Treatment × time	$F(177) = 8.71, p < .001$.10
	Treatment × group	$F(172) = 3.25, p < .05$.04
TCK	Group	$F(168) = 5.22, p < .01$.06
	Treatment × time	$F(150) = 7.10, p < .01$.09
PCK	Treatment × time	$F(137) = 18.01, p < .001$.21
	Treatment × group	$F(149) = 5.66, p < .01$.09
TPCK	Treatment × time	$F(153) = 19.16, p < .001$.20
	Treatment × group	$F(172) = 3.25, p < .05$.06

DV dependent variable, LMM linear mixed model

Only statistically significant effects ($p < .05$) are shown; for interaction effects, main effects for factors involved in the interaction are not shown

of intervention. For five of the seven scales (TPCK, PK, TPK, PCK, CK), there were significant interaction effects time \times treatment and time \times group of mainly moderate to strong power indicating that the progression of self-reported TPACK developed differently over time depending on treatment received and differently between the four study groups. For most scales, the effect size was stronger for the time \times treatment interaction than for the time \times group interaction. That is, the main driver for TPACK progression in this study is about receiving the treatment and not belonging to a certain study group. When comparing the two treatments (IG and CG), we found that for post-tests, students in the IG significantly reported higher agreement than students in the CG concerning the statements measuring TPCK, TPK, and TCK, but not CK, PK, TK, and PCK.

Discussion

Integrating digital tools into science teacher education presents an emerging field of science education research (Angeli & Valanides, 2015; Herring et al., 2016; Huang et al., 2022). The present study adds to the literature the development and application of a domain-specific questionnaire that allows investigation of how pre-service teachers' TPACK developed over 13 weeks of a regular teacher education program. The need for such a domain-specific questionnaire derives from science education research that aims to understand how subject-specific university-based teacher education contributes to professional development. In this regard, our efforts align with research on PCK, in which specific instruments have been developed for subject-specific domains (Großschedl et al., 2018, 2019). Hence, the development of longitudinal TPACK in specific domains can be supported by questionnaires that are specific to that domain. However, such questionnaires are scarce. The present study stands in to fill this gap in the domain of applying digital tools for teaching human biology in the context of a biology teacher education program. We discuss the findings of our empirical study with respect to (1) issues related to the reliability and validity of the questionnaire, (2) findings on TPACK professional development over the 13 weeks of intervention, and (3) the advantages of implementing a sequential application of the questionnaire through intermediate testing.

Issues Related to the Reliability and Validity of the Questionnaire

For questionnaire development, we sourced items from the TPACK.xs instrument by Schmid et al. (2020) and systematically rephrased the items for the domain of human biology without changing the nature of the original items.

In doing so, we succeeded in forming reliable scales with strong item-discriminatory power and internal consistency for all seven TPACK scales (see Table 3; Kline, 2016). The good psychometric properties of the questionnaire allowed us to identify differences in the TPACK development (Taber, 2018). Furthermore, we were able to maintain some of the benefits of the original TPCK.xs questionnaire, such as a short administration time (Schmid et al., 2020).

However, the procedure of questionnaire development is not new. It has been successfully applied in other areas of questionnaire development for science education research (e.g., Nature of Science, Voitle et al., 2022). In previous TPACK studies, item adaptation focused on rephrasing general contexts ("in my teaching," "in my subject"; Schmid et al., 2020) with subject-specific framings (i.e., "deal with biological issues," "for learning biology with technology"; Mahler & Arnold, 2022; von Kotzebue, 2022b). In the present study, we went one step further and adapted the items to be domain specific. This provides a basis for assessing self-reported knowledge in a specific domain where specific digital tools are used. Hence, we accounted for the contextual dependence of teachers' technology use (Mishra, 2019). That is, teachers may be familiar with digital tools dedicated to subject-matter learning in one domain but not in another (contextual knowledge at the microlevel; Brianza et al., 2022). We suggest that the procedure of item developing can be transferred to other subject-specific domains of TPCK research, where researchers are interested in investigating domain-specific TPCK development (Rosenberg & Koehler, 2015; Wang et al., 2018).

The findings concerning the factorial validity of the TPACK scales revealed that a 7-factorial model fit better to the data than models for the transformative and integrative perspectives of TPCK (Scherer et al., 2017; Schmid et al., 2020; von Kotzebue, 2022b). While a 7-factorial model has similarly been described as the best-fitting model in previous studies with biology student teachers (e.g., Mahler & Arnold, 2022), findings on the transformative vs. integrative view came to different conclusions. Von Kotzebue et al. (2022b) found in her study with Austrian university students that the transformative view tended to fit better to the data. In contrast to von Kotzebue (2022b), who reports findings only for the technology-related TPACK constructs (TK, TCK, TPK, TPCK), we took a broader perspective and further added the non-technology-related TPACK constructs to the model (PK, CK, PCK). Based on latent variable modeling, we found that TK was significantly correlated with TCK and TPK while TK was also significantly correlated with CK, PK, and PCK, but to a lesser extent (see Table 4). To us, these findings mean that technology-related knowledge ("T" in TPACK) is not acquired in isolation from the contextual background (Mishra, 2019). That is, when TPACK is assessed in the context of a teacher education

program, a specific domain (e.g., human biology) should be considered to gain a more comprehensive view of knowledge acquisition and development (Tan et al., 2022). Further research is needed to address to what extent TPACK should be viewed as a generic (i.e., knowledge for general context) and how the generic nature influences the domain-specific TPACK development (Lachner et al., 2021). Moreover, the need for reflecting a specific context in which knowledge is acquired helps to acknowledge that teachers' professional knowledge has a strongly contextualized nature (Brianza et al., 2022; Putnam & Borko, 2000).

Findings on TPACK Professional Development over the 13 Weeks of Intervention

As a main finding, LMM analyses helped identify the progression of teachers' TPACK over the 13 weeks of intervention. The development took place in both non-technology-related constructs (e.g., PK and CK) and technology-related constructs (e.g., TCK and TPACK). We expected progression in the non-technology-related constructs, since the two treatments were part of a regular teacher education program that typically aims to convey pedagogical knowledge and content knowledge (Huang et al., 2022). Progression in technology-related knowledge was found to be a function of the treatment received and was stronger in the intervention groups than in the control groups. That is, working with anatomy apps on a regular basis affects the development of the TPACK. Students who used the anatomy app during their studies outperformed students in the control groups regarding TPACK, TPK, and TCK. Thus, this study confirms that actively working with digital tools in the context of human biology can make a significant contribution to the professional knowledge growth of students during teacher education programs (Huang et al., 2022; Wang et al., 2018).

As an interesting finding, students enrolled in the master's study program (study group 4) did not perform significantly better than students enrolled in the initial study program (study groups 1 to 3). Based on findings from prior PCK research, we would expect stronger differences since a higher number of semesters attended in higher education correspond with a more developed PCK (Kleickmann et al., 2013). We considered these differences by applying a linear mixed model with fixed and random effects and checked for the effects of the study groups (Kuznetsova et al., 2017).

These results can be interpreted in different ways. First, when considering the control groups (groups 2 and 3) in comparison to intervention group 4, it appears that the students in group 4, who had more experience with educational training in the form of science education seminars visited and credit points awarded, may have been more self-critical when assessing their TPACK development. This led to lower scores (underestimation)

compared to the students in study groups 1 to 3, who had less experience and, therefore, a possible less critical view of their TPACK development, potentially resulting in higher scores (overestimation). This interpretation is consistent with the findings of Max et al. (2022), who discovered that low-performing students tend to overestimate their TPACK more than high-performing students do, as demonstrated by a comparison of self-assessment questionnaires and performance-based assessments. In contrast, when analyzing the findings of two study groups at the same proficiency level (intervention group 1 and control groups 2 and 3, bachelor study program), it was discovered that students in the intervention groups reported stronger development in TPACK from pre-test to post-test. This leads us to infer that continuous work with digital anatomy tools leads to a higher score on the self-assessment of TPACK, thus providing evidence for the usefulness of digital anatomy tools to support professional development.

Advantages of Integrating a Sequential Application of the Questionnaire Through Intermediate Testing

Applying the questionnaire at three time points (pre-test, intermediate test, and post-test) was characteristic of two study groups (groups 1 and 2). This was made possible based on the relatively short time needed to administer the questionnaire (Schmid et al., 2020). Implementing an intermediate test at week five allowed us to investigate sequential TPACK development in more detail. The analysis uncovered that there was hardly any difference in TPACK between study group 1 (intervention group) and study group 2 (control group) between the pre- and intermediate test. That is, in intervention group 1, we did not measure TPACK development beyond the effect of implementing the same questionnaire twice (test-retest effect; Kline, 2016). In contrast, when comparing intermediate to post-test development, solely study group 1 reported a strong increase in TPACK and there is a lack of further professional development in study group 2. This "boost effect" for study group 1 following the intermediate test can be attributed due to the fact that the students in study group 1 hold presentations based on the 3D anatomy apps and digital tools that were used to support animal specimen dissection whereas students in study group 2 used traditional digital presentation methods and accessed traditional anatomical models. These students took full advantage of the potential of using digital tools in human anatomy (Chakraborty & Cooperstein, 2018; Lewis et al., 2014; Moro et al., 2017). By implementing the intermediate test, we were able to shed light on the sequential TPACK development that occurred. Typically, sequential developments are hidden when researchers solely focus on pre-post-comparisons (Hofer & Grandgenett, 2014). Based on the findings of the present study, we advocate applying

TPACK questionnaires sequentially, since they allow for a more fine-grained analysis of TPACK professional development. However, implementing three time points may not be fine-grained enough since we do not know the exact time points where the boost effect in the intervention groups and the plateau and decline in the control groups start.

Limitations

In the present study, we relied on students' self-reported data, which are typically used when investigating the development of TPACK in science education (Schmid et al., 2021). However, this procedure is often criticized for being too subjective and not measuring knowledge (Krauskopf & Forssell, 2018). Thus, the present study is limited, in that when referring to the development of TPACK, how the students themselves perceived their progression of knowledge. Alternative methods for evaluating TPACK have focused on performance-based assessments (e.g., observing the teaching practices of teachers) and assessing the quality of teaching instruction in relation to the TPACK framework, such as von Kotzebue (2022b) and Max et al. (2022). However, in the present study, we did not have the means to apply a performance-based assessment of TPACK, potentially resulting in overestimation and underestimation of students' TPACK.

Furthermore, due to the federal system of teacher education and the involvement of study groups from different parts of Germany, the time dedicated to teacher education courses differed (see Table 1). While we acknowledge that the data analyzed in this study stems from in situ university courses and not from a laboratory setting, this characteristic also posits a limitation. These findings should be interpreted with caution, as different study groups spent different times in class. We still evaluated the study population as valuable because two pairs of study groups (groups 1 and 2, as well as groups 3 and 4) had mostly similar settings regarding course time. This study may have overlooked variables other than those described in the sample description that could have impacted the data and were not accounted for in the analysis.

Conclusions and Further Directions

The present study investigated how using digital tools (e.g., 3D anatomy apps) influenced the development of TPACK in a biology student teacher population over a course of 13 weeks. This study contributes to technology-related science education research by developing a domain-specific TPACK questionnaire. Applying a quasi-experimental study

design and following a systematic approach to rephrase the existing TPACK.xs questionnaire by Schmid et al. (2020) enabled us to develop a questionnaire that allows for reliable measurement and valid interpretation of TPACK. The integration of intermediate data collection led to a fine-grained analysis of sequential TPACK development and posited a huge advantage of the questionnaire. The sequential analyses highlighted that working with digital tools on a regular basis in the intervention groups could boost TPACK development. Going forward, we aim to investigate how TPACK development corresponds to a specific usage of digital tools (i.e., usage that facilitates or hinders TPACK development). To gain a more comprehensive picture of teachers' professional development, mixed-method surveys (Tondeur et al., 2020) and studying teachers' technology-related beliefs (Ertmer et al., 2015; von Kotzebue, 2022a) may be a promising approach.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10956-024-10108-w>.

Acknowledgements We want to thank Simon Blauza, Leonie Schrader, Marie Neumann, Sebastian Linnartz, Andrea Ocklitz, Annette Philipps, and Hanna Parkhomhuck for their generous support with data collection.

Funding Open Access funding enabled and organized by Projekt DEAL. This study received support from the Joachim Herz Foundation, Hamburg, Germany (Grant Number 850041).

Data Availability The data that support the findings of this study are available from the corresponding author upon request.

Declarations

Ethics Approval and Consent to Participate The study was performed in accordance with the ethical standards of the 1964 Declaration of Helsinki. All participants were informed about the data protection measures undertaken to guarantee anonymity, with accordance to GDPR, and gave informed consent. Participation in the study could be stopped at any time without consequences. No incentives were given.

Conflict of Interest The last author, Benedikt Heuckmann, was employed at the same university (Institute for Science Education, Leibniz University Hannover) as the first author, Leona Mölgen, during the time the study was conducted. Other than that, the authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abbitt, J. T. (2014). Measuring technological pedagogical content knowledge in preservice teacher education. *Journal of Research on Technology in Education*, 43(4), 281–300. <https://doi.org/10.1080/15391523.2011.10782573>
- Angeli, C., & Valanides, N. (2009). Epistemological and methodological issues for the conceptualization, development, and assessment of ICT-TPCK: Advances in technological pedagogical content knowledge (TPCK). *Computers & Education*, 52(1), 154–168. <https://doi.org/10.1016/J.COMPEDU.2008.07.006>
- Angeli, C., & Valanides, N. (Eds.) (2015). *Technological Pedagogical Content Knowledge: Exploring, Developing, and Assessing TPCK*. Springer. <https://doi.org/10.1007/978-1-4899-8080-9>
- Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Journal of Psychology*, 223(1), 3–13. <https://doi.org/10.1027/2151-2604/a000194>
- Brianza, E., Schmid, M., Tondeur, J., & Petko, D. (2022). Situating TPCK: A systematic literature review of context as a domain of knowledge. *Contemporary Issues in Technology and Teacher Education*, 22(4), 707–753.
- Chai, C. S. (2019). Teacher professional development for science, technology, engineering and mathematics (STEM) education: A review from the perspectives of technological pedagogical content (TPACK). *Asia-Pacific Education Researcher*, 28(1), 5–13. <https://doi.org/10.1007/S40299-018-0400-7/TABLES/1>
- Chai, C. S., Hwee, J., Koh, L., & Tsai, C.-C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society*, 16(2), 31–51. <https://doi.org/10.2307/jeductechsoci.16.2.31>
- Chakraborty, T. R., & Cooperstein, D. F. (2018). Exploring anatomy and physiology using iPad applications. *Anatomical Sciences Education*, 11(4), 336–345. <https://doi.org/10.1002/ase.1747>
- Cheng, K. H., & Tsai, C. C. (2013). Affordances of augmented reality in science learning: Suggestions for future research. *Journal of Science Education and Technology*, 22(4), 449–462. <https://doi.org/10.1007/S10956-012-9405-9>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159.
- Doering, A., Veletsianos, G., Scharber, C., & Miller, C. (2009). Using the technological, pedagogical, and content knowledge framework to design online learning environments and professional development. *Journal of Educational Computing Research*, 41(3), 319–346. <https://doi.org/10.2190/EC.41.3.D>
- Ertmer, P. A., Ottenbreit-Leftwich, A. T., & Tondeur, J. (2015). Teachers' beliefs and uses of technology to support 21st-century teaching and learning. In H. Fives & M. G. Gill (Eds.), *International Handbook of Research on Teachers' Beliefs* (pp. 403–418). Routledge.
- European Commission. (2020). *Digital education action plan (2021–2027). Resetting education and training for the digital age*. <https://education.ec.europa.eu/focus-topics/digital-education/action-plan>
- Großschedl, J., Mahler, D., & Harms, U. (2018). Construction and evaluation of an instrument to measure content knowledge in biology: The CK-IBI. *Education Sciences*, 8(3), 145. <https://doi.org/10.3390/EDUCSCI8030145>
- Großschedl, J., Welter, V., & Harms, U. (2019). A new instrument for measuring pre-service biology teachers' pedagogical content knowledge: The PCK-IBI. *Journal of Research in Science Teaching*, 56(4), 402–439. <https://doi.org/10.1002/TEA.21482>
- Haglund, J., Jeppsson, F., & Schönborn, K. J. (Eds.). (2022). *Thermal cameras in science education*. Springer. <https://doi.org/10.1007/978-3-030-85288-7>
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends*, 60(5), 433–441. <https://doi.org/10.1007/S11528-016-0091-Y/TABLES/1>
- Herring, M. C., Koehler, M. J., & Mishra, P. (2016). *Handbook of Technological Pedagogical Content Knowledge (TPACK) for Educators* (2nd Edition). Routledge. <https://doi.org/10.4324/9781315771328>
- Hillmayr, D., Zierenwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/J.COMPEDU.2020.103897>
- Hofer, M., & Grandgenett, N. (2014). TPCK development in teacher education A longitudinal study of preservice teachers in a secondary M.A.Ed. Program. *Journal of Research on Technology in Education*, 45(1), 83–106. <https://doi.org/10.1080/15391523.2012.10782598>
- Hoyle, R. H. (2012). *Handbook of structural equation modeling*. Guilford Press.
- Huang, B., Siu-Yung Jong, M., Tu, Y. F., Hwang, G. J., Chai, C. S., & Yi-Chao Jiang, M. (2022). Trends and exemplary practices of STEM teacher professional development programs in K-12 contexts: A systematic review of empirical studies. *Computers & Education*, 189, 104577. <https://doi.org/10.1016/J.COMPEDU.2022.104577>
- Hume, A., Cooper, R., & Borowski, A. (2019). *Repositioning Pedagogical Content Knowledge in Teachers' Knowledge for Teaching Science*. Springer. <https://doi.org/10.1007/978-981-13-5898-2>
- Kleickmann, T., Richter, D., Kunter, M., Elsnar, J., Besser, M., Krauss, S., & Baumert, J. (2013). Teachers' content knowledge and pedagogical content knowledge: The role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106. <https://doi.org/10.1177/0022487112460398>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling. Methodology in the social sciences* (4th ed.). Guilford Press.
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Krauskopf, K., & Forsell, K. (2018). When knowing is believing: A multi-trait analysis of self-reported TPCK. *Journal of Computer Assisted Learning*, 34(5), 482–491. <https://doi.org/10.1111/JCAL.12253>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/JSS.V082.I13>
- Lachner, A., Fabian, A., Franke, U., Preiß, J., Jacob, L., Führer, C., Küchler, U., Paravicini, W., Randler, C., & Thomas, P. (2021). Fostering pre-service teachers' technological pedagogical content knowledge (TPACK): A quasi-experimental field study. *Computers & Education*, 174, 104304. <https://doi.org/10.1016/J.COMPEDU.2021.104304>
- Lewis, T. L., Burnett, B., Tunstall, R. G., & Abrahams, P. H. (2014). Complementing anatomy education using three-dimensional anatomy mobile software applications on tablet computers. *Clinical Anatomy*, 27(3), 313–320. <https://doi.org/10.1002/CA.22256>
- Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, 95, 174–187. <https://doi.org/10.1016/J.COMPEDU.2016.01.001>
- Mahler, D., & Arnold, J. (2022). MaSter-Bio – Messinstrument für das akademische Selbstkonzept zum technologiebezogenen Professionswissen von angehenden Biologielehrpersonen. *Zeitschrift für Didaktik der Naturwissenschaften*, 28(1), 1–16. <https://doi.org/10.1007/S40573-022-00137-6>
- Max, A. L., Lukas, S., & Weitzel, H. (2022). The relationship between self-assessment and performance in learning TPCK: Are self-assessments a good way to support preservice teachers' learning?

- Journal of Computer Assisted Learning*, 38(4), 1160–1172. <https://doi.org/10.1111/JCAL.12674>
- Mishra, P. (2019). Considering contextual knowledge: The TPACK diagram gets an upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, 10(6), 549–559. <https://doi.org/10.1002/ASE.1696>
- Niess, M. L., van Zee, E. H., & Gillow-Wiles, H. (2014). Knowledge growth in teaching mathematics/science with spreadsheets. *Journal of Digital Learning in Teacher Education*, 27(2), 42–52. <https://doi.org/10.1080/21532974.2010.10784657>
- Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, 29(1), 4–15. https://doi.org/10.3102/0013189X029001004/ASSET/0013189X029001004.FP.PNG_V03
- R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rienties, B., Brouwer, N., & Lygo-Baker, S. (2013). The effects of online professional development on higher education teachers' beliefs and intentions towards learning facilitation and technology. *Teaching and Teacher Education*, 29(1), 122–131. <https://doi.org/10.1016/j.tate.2012.09.002>
- Robitzsch, A. (2020). Why ordinal variables can (almost) always be treated as continuous variables: Clarifying assumptions of robust continuous and ordinal factor analysis estimation methods. *Frontiers in Education*, 5, 177. <https://doi.org/10.3389/FEDUC.2020.589965/BIBTEX>
- Rosenberg, J. M., & Koehler, M. J. (2015). Context and technological pedagogical content knowledge (TPACK): A systematic review. *Journal of Research on Technology in Education*, 47(3), 186–210. <https://doi.org/10.1080/105391523.2015.1052663>
- Rossee, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Scherer, R., Tondeur, J., & Siddiq, F. (2017). On the quest for validity: Testing the factor structure and measurement invariance of the technology-dimensions in the Technological, Pedagogical, and Content Knowledge (TPACK) model. *Computers & Education*, 112, 1–17. <https://doi.org/10.1016/j.compedu.2017.04.012>
- Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACKxs) 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>
- Schmidt, D. A., Baran, E., Thompson, A. D., Mishra, P., Koehler, M. J., & Shin, T. S. (2009). Technological pedagogical content knowledge (TPACK). *Journal of Research on Technology in Education*, 42(2), 123–149. <https://doi.org/10.1080/105391523.2009.10782544>
- Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Sekretariat der Kultusministerkonferenz [KMK, 2019]. [Secretary of the Conference of the Ministers of Education and Cultural Affairs]. (2019). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Beschluss der Kultusministerkonferenz vom 16.10.2008 i. d. F. vom 16.05.2019*. [Joint national standards for subject-matter education in teacher education].
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/S11165-016-9602-2/TABLES/1>
- Tan, S. C., Teo, T. W., & Tsai, C.-C. (2022). Digital technologies and professional learning of science teachers: A technological pedagogical content knowledge (TPACK) perspective. In J. Luft & M. G. Jones (Eds.), *Handbook of research on science teacher education* (pp. 313–324). Routledge. <https://doi.org/10.4324/9781003098478-27>
- Thoms, L.-J., Hoyer, C., & Girwidz, R. (2022). A teacher training course on using digital media for acquisition, visualization and 3D printing of complex data and for fostering pupils' experimental skills. In J. B. Marks, P. Galea, S. Gatt, & D. Sands (Eds.), *Physics teacher education. What matters?* (pp. 75–90). Springer. https://doi.org/10.1007/978-3-031-06193-6_6
- Tondeur, J., Scherer, R., Siddiq, F., & Baran, E. (2020). Enhancing pre-service teachers' technological pedagogical content knowledge (TPACK): A mixed-method study. *Educational Technology Research and Development*, 68(1), 319–343. <https://doi.org/10.1007/S11423-019-09692-1/TABLES/4>
- UNESCO. (2018). UNESCO ICT competency framework for teachers. United Nations Educational, Scientific and Cultural Organization. Retrieved March 19, 2024, from <https://unesdoc.unesco.org/ark:/48223/pf00000265721>
- Voitle, F., Heuckmann, B., Kampa, N., & Kremer, K. (2022). Assessing students' epistemic beliefs related to professional and school science. *International Journal of Science Education*, 44(6), 1000–1020. <https://doi.org/10.1080/09500693.2022.2059821>
- von Kotzebue, L. (2022a). Beliefs, self-reported or performance-assessed TPACK: What can predict the quality of technology-enhanced biology lesson plans? *Journal of Science Education and Technology*, 31, 570–582. <https://doi.org/10.1007/s10956-022-09974-z>
- von Kotzebue, L. (2022b). Two is better than one—Examining biology-specific TPACK and its T-dimensions from two angles. *Journal of Research on Technology in Education*. <https://doi.org/10.1080/15391523.2022.2030268>
- von Kotzebue, L., Meier, M., Finger, A., Kremser, E., Huwer, J., Thoms, L. J., Becker, S., Bruckermann, T., & Thyssen, C. (2021). The framework DiKoLAN (digital competencies for teaching in science education) as basis for the self-assessment tool DiKoLAN-Grid. *Education Sciences*, 11(12), 775. <https://doi.org/10.3390/EDUCSCI11120775>
- Wang, W., Schmidt-Crawford, D., & Jin, Y. (2018). Preservice teachers' TPACK development: A review of literature. *Journal of Digital Learning in Teacher Education*, 34(4), 234–258. <https://doi.org/10.1080/21532974.2018.1498039>
- Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α Revelle's β and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, 70(1), 123–133. <https://doi.org/10.1007/s11336-003-0974-7>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.