Context-specificity to reduce bias in self-assessments: Comparing teachers' scenario-based self-assessment and objective assessment of technological knowledge

Tamara Kastorff[a], Michael Sailer[a,b], Johanna Vejvoda[a], Florian Schultz-Pernice[a], Valentin Hartmann[a], Anna Hertl[a], Sonja Berger[a] and Karsten Stegmann[a]

[a]Ludwig-Maximilians-Universität München, Munich, Germany; [b]Friedrich-Alexander Universität, Erlangen, Germany

Introduction

The COVID-19 pandemic has severely challenged teachers all over the world. Teachers were required to set up online learning environments and video lessons for distance education, a challenge for which sound technological knowledge (TK) is helpful (Koehler & Mishra, 2009). As the use of technology in teaching and learning processes has been inevitable during the COVID-19 pandemic, teachers' TK became a key requirement for teaching and learning. TK refers to the teachers' knowledge of technologies, such as digital tools and educational technologies (Lachner et al., 2019, p. 7).

Numerous studies have examined teachers' TK by using self-assessment measures, for which teachers rate or indicate how confident they feel about their skills and knowledge (e.g., Mourlam et al., 2021; Rubach & Lazarides, 2021; Schmid et al., 2020). Self-assessments make a valuable contribution to educational research because they are closely related how teachers intend to use technology in their classrooms (Scherer et al., 2015) and further provide rich and meaningful information for educational researchers without much effort (Seufert et al., 2021). Nonetheless, the validity of self-assessments has been criticized. Several authors (Aesaert et al., 2017; Hatlevik et al., 2018; Scherer et al., 2017) have argued that self-assessments capture individuals' self-perceived abilities, knowledge, or skills, which might not be consistent with their *actual* performance. In line with this argument, studies have regularly shown only small linear relationships between

CONTACT Tamara Kastorff tamara.kastorff@psy.lmu.de Department of Psychology, Ludwig-Maximilians-Universität München, Leopoldstraße 15, Munich 80802, Germany

self-assessments and objective assessments of teachers' TK (e.g., Akyuz, 2018; Baier & Kunter, 2020; Drummond & Sweeney, 2017), raising questions about whether teachers' perceived ability matches their actual knowledge and skills, and thus establishing a need for more objective assessments to be implemented in educational research (Lachner et al., 2019; Petko, 2020).

However, objective performance tests are difficult to administer because participants are often reluctant to participate in objective performance tests (Kleinert et al., 2015) and objective measurement instruments often address very specific uses of technology, such as asking participants if they can operate a particular program (Petko, 2020). Very specific uses of technology might not provide a comprehensive perspective on the successful use of technology in the sense of the level of technology use that is necessary to function in an information-based society (Fraillon et al., 2020). Further, objective assessment measures of TK can quickly become outdated due to advances in technology (Siddiq et al., 2016).

Because objective assessments are often poorly accepted by participants in practice, and self-assessments suffer from biases due to low correlations between self-assessment and objective assessment measures (e.g., Drummond & Sweeney, 2017), Sailer et al. (2021) attempted to make self-assessments more accurate. To do so, the authors proposed scenario-based self-assessments to improve the accuracy of self-assessment measures by providing concrete scenarios. These scenarios confront participants with detailed information in a context-specific situation to address the problem of potential self-assessment biases, such as social desirability or ambiguity. The relevance of context-specificity in self-assessments has also been demonstrated in previous research. Talsma et al. (2018) found evidence that more context-specific self-assessment measures have stronger correlations with objective performance measures than less context-specific self-assessment measures do. Thus, scenarios in self-assessments represent a concrete problem-solving scenario that focuses participants' general ability expectations on an example case so that participants can better visualize what the situation or task at hand has at stake (Sailer et al., 2021).

Given the overt relevance of teachers' TK in times of digitalization and remote learning, and given the current call to identify potential biases in self-assessments in educational science, the main purpose of this study is to analyze the relationship between scenario-based self-assessments and objective assessments of teachers' TK.

In summary, teachers' TK is of crucial importance in educational science in times of digitalization and remote learning. However, objective assessment measures are often difficult to implement, particularly concerning teachers' use of technology. This study aims to investigate the extent to which scenario-based self-assessments of teachers' TK can serve as anchors by helping them to identify the relevant skills that are required in concrete and authentic situations (Sailer et al., 2021), and might be therefore effective to bring objective and subjective measures closer together (Van Soest et al., 2011).

Technological knowledge (TK)

Regarding the integration of technology into teaching and learning processes, Koehler and Mishra (2009) proposed the TPACK framework for educational research. The authors established the TPACK framework based on Shulman's (1986) constructs of pedagogical content knowledge (PCK) and extended PCK by adding technology (T). Accordingly, the TPACK framework describes the relationships between the constructs of content, pedagogy, and technology. Within the TPACK framework, TK is an essential component of the successful implementation of technology into teaching and learning processes. Although TK is a central component within the TPACK framework, empirical studies regarding the factor structure of the TPACK framework have shown a different picture. For example, both Scherer et al. (2017) and Lachner et al. (2019) identified TK as primarily independent of the other facets of the TPACK framework, and consequently, the authors assumed that TK might be independent of teaching and learning processes.

Koehler and Mishra (2009) broadly defined TK as "understanding information technology broadly enough to apply it productively at work and in everyday life, recognizing when information technology can assist or impede the achievement of a goal, and continually adapting to changes in information technology" (p. 64). Against the relevance of teachers' TK, the Standing Conference of the Ministers of Education and Cultural Affairs (KMK) published a strategy for education in a digital world (Kultusministerkonferenz (KMK), 2019a). The Standing Conference of the Ministers of Education and Cultural Affairs (KMK) is an association responsible for education and schools, vocational training, and research (Kultusministerkonferenz (KMK), 2019b). Therefore, the KMK plays an important role in Germany as an instrument for the coordination and development of education, including in the area of digitization and the digital skills of students and teachers. According to the KMK (2019a), TK is an essential part of a person's ability to participate successfully in society and later in professional life and is broadly described from a holistic perspective as the *basic digital skills* needed to use technologies to collect, manage, produce, and exchange information. Further, according to the KMK (2019a), basic digital skills encompass core skills concerning the use of technology, such as communicating and collaborating with technology. Such skills are an integral part of teacher education and training (Sailer et al., 2021; Forschungsgruppe Lehrerbildung Digitaler Campus Bayern [Research Group Teacher Education Digital Campus Bavaria], 2017).

The basic digital skills can be narrowed down to five core concepts in using technology (KMK, 2019a):

1. *Operating and applying technology*. According to the KMK (2019a), operating and applying technology refers to the skills and knowledge needed to operate and apply technology appropriately and purposefully, including knowing the basic principles and functions of technology. In addition, one should be able to optimize one's own use of technology.
2. *Searching for and processing information with technology*. Searching for and processing information with technology incorporates the development of search strategies to obtain the desired information and to purposefully select the appropriate information. It also includes the ability to store, summarize, structure, and critically evaluate information in a purposeful manner.
3. *Communicating and collaborating with technology*. Collaborating and Communicating with technology describes the ability to use technology to collaborate and communicate with others, a skill that is necessary to successfully participate in society. Collaborating and communicating with technology must also consider rules of engagement and others' personal rights.
4. *Producing and presenting information with technology*. According to the KMK (2019a), producing and presenting information with technology refers to selecting and using media products appropriately and considering formal design features and intentions to present adequate information with technology.
5. *Analyzing and reflecting information with technology*. Analyzing and reflecting describes the competence to analyze and evaluate the content, design features, and structure of media products, whereby the interest-driven dissemination of media content is recognized and critically evaluated. In addition, one should be able to assess the potential and risks of using media for oneself and society.

The five core components presented by the KMK (2019a) are also well-documented and established in widely accepted theoretical frameworks on basic digital skills, such as the Digital Competence of Educators framework (Ferrari, 2013) or the ICILS framework (Fraillon et al., 2020). In line with the KMK (2019a), the DigCompEdu framework targets the specific basic digital skills of educators as prerequisites for the ability to facilitate student learning (Siddiq et al., 2016).

Consequently, it can be concluded that there are currently two major ways to operationalize TK in educational science. On the one hand, TK is currently assessed from a functional perspective (see Lachner et al., 2019; Senkbeil et al., 2013), including multiple facets concerning the operational use of technology. On the other hand, TK has recently been represented by broader approaches, such as the KMK (2019a), which describes a more holistic use of technology based on five core concepts that people must meet to successfully participate in an information-based society. In the context of this study, we use the terms *TK* and *digital skills* interchangeably, thereby suggesting that people must incorporate qualities from both the technical operational skills and the five core concepts derived from the KMK framework (2019) in order to successfully participate in an information-based society.

Self-assessment and objective assessment

Whereas there is a large body of research on how teachers self-assess their skills (e.g., Lucas et al., 2021; Scherer et al., 2017), less attention has been paid to teachers' actual performance on objective assessment measures (Petko, 2020; Seufert et al., 2021). In current research practice, teachers' skills are often assessed with self-reports in which teachers indicate how confident they feel about whether they have a particular skill (e.g., Schmid et al., 2021).

Nevertheless, it is important to emphasize that self-assessment is not an accurate measure of an individual's objective performance (Hatlevik et al., 2018). In line with this, a common issue when assessing skills via self-assessment is that estimations rely on individuals' ability to self-assess accurately. This can be problematic because the ability of individuals to assess their performance validly must be regarded as very heterogeneous (Van Soest et al., 2011; van Vliet et al., 1994). Thus, the relationship between self-assessment and objective assessment measures is often weak (Akyuz, 2018; Drummond & Sweeney, 2017; Dunning et al., 2004).

Self-assessment measures are also widely applied to measure the construct of *self-efficacy* (Bandura, 1977). Research on self-efficacy has shown that the effectiveness of self-efficacy measures in educational research is highly context-specific (Bandura, 1977; Rohatgi et al., 2016), which suggests that the relevance of context-specificity also applies to self-assessment measures (Sailer et al., 2021). Moreover, research (e.g., Bandura, 1977) has shown that it is important to distinguish between *general* self-assessment and *domain-specific* self-assessment, which can be described as domain-specific self-efficacy and general self-efficacy. General self-assessment or self-efficacy refer to overarching, general situations (e.g., "I am good at what I do"; Rohatgi et al., 2016), and domain-specific self-efficacy refers to concrete domains, such as Information Communication and Technology (ICT) self-efficacy (e.g., "I can type terms correctly into search engines"). Scholars already agree that domain-specific self-assessment or self-efficacy measures are more suitable for making concrete statements about a person's self-efficacy in a specific domain than about a person's general self-efficacy (Rohatgi et al., 2016). Moreover, Scherer and Siddiq (2015) showed that teachers' domain-specific information and communication technology (ICT) self-efficacy further consists of three separate constructs: basic operational skills, advanced operational and collaborative skills, and the use of technology for instructional purposes. According to Hatlevik and Hatlevik (2018), Scherer and Siddiq (2015) findings can be interpreted to mean that domain-specific ICT self-efficacy measures still leave too much room for heterogeneous interpretations of specific tasks related to ICT self-efficacy. Therefore, despite the agreement that domain-specific self-efficacy is more suitable for making concrete statements about a person's self-efficacy, domain-specific self-efficacy (e.g., ICT self-efficacy) might still be too general with regard to context-sensitivity.

The relevance of context-specificity in self-assessment was also supported by the meta-analysis by Talsma et al. (2018) on self-efficacy and academic performance. Accordingly, the authors reported that the more context-specific self-assessment measures are, the more strongly they are correlated with objective assessment measures than less context-specific self-assessment measures. Further, according to Peura et al. (2019), context-specific self-assessment measures have stronger relationships with objective assessment measures than less context-specific self-assessment measures, although few studies have examined this relationship to date. Furthermore, according to

Scheiter (2021), item formulations in current self-assessment instruments are often too vague to make concrete statements about a person's knowledge or skills. According to the author, the vague item formulations in self-assessments, such as "I can use digital media efficiently in the classroom," leave participants too much room for interpretation because teachers tend to orient themselves more toward theoretical conventions of high-quality teaching with digital media than toward actual skills.

Scheiter (2021) postulated that vignette-based items would be a possible solution as they guide teachers to represent their own skills in an action-oriented and situation-based test format, which is an improvement over the usually vague and therefore problematic item formulations used in regular self-assessments. In line with King et al. (2004) and King and Wand (2007) findings, short vignettes can correct for respondents' different understandings of scenarios given that regular self-assessments often lack contextual information, thus resulting in different respondents understanding the same self-assessment question in different ways. Van Soest et al. (2011) were able to validate this assumption empirically: according to the authors, concrete scenarios such as described in vignettes, are suitable in bringing self-assessments in line with objective assessment. Further, on the basis of King et al. (2004) assumptions, Sailer et al. (2021) developed a similar approach with their scenario-based assessments of teachers' digital skills. In Sailer et al.'s (2021) study, participants were placed in a concrete scenario where they were asked to assess their knowledge and skills with respect to a particular scenario, such as the extent to which teachers feel able to implement group work with tablets in the classroom. According to the authors, compared with regular self-assessments, the concrete formulation of a scenario acts as an "anchor" that can help teachers assess their own skills and attitudes more accurately and validly (p. 7).

In summary, scenario-based self-assessment might be a suitable aid for participants to evaluate their skills more accurately in situations in which it is crucial to apply the corresponding skills.

The present study

Given the importance of teachers' TK, particularly considering the COVID-19 pandemic and regarding current calls to become aware of potential biases concerning self-assessment, we aim to gain more insight into the relationship between teachers' scenario-based self-assessments and objective assessments, focusing on TK. To do so, we pose the following research question:

> RQ: How can (student) teachers' objectively assessed TK be predicted by scenario-based self-assessments of their TK?

Objective assessment measures are often difficult to implement because subjects are often reluctant to participate in objective assessment studies (Kleinert et al., 2015) and objective assessment measures, especially with regard to TK, are either very specific (Petko, 2020) or quickly become outdated (Siddiq et al., 2016). Therefore, self-assessment measures are often used in educational research to gain information regarding teachers' digital skills. However, it has been shown that the relationships between self-assessment measures and objective assessment measures related to teachers' digital skills are often non substantial (see Baier & Kunter, 2020; Drummond & Sweeney, 2017), suggesting that individuals' ability to accurately assess their performance must be considered heterogeneous (van Vliet et al., 1994). Van Soest et al. (2011) provided evidence that concrete scenarios are suitable for matching self-assessments with objective assessment measures related to personal behavior. Consistent with this, scholars in educational science (Peura et al., 2019; Rohatgi et al., 2016; Scheiter, 2021; Talsma et al., 2018) postulated the assumption that a clear framework or reference, as realized in scenario-based self-assessment, could provide a clearer context for individuals to assess their actual skills and approximate the results of objective assessment measures more accurately. In line with this assumption, the present study explores the extent to which teachers' scenario-based self-assessment of TK aligns with the results of teachers' objective measurement of TK.

Method

Participants and procedure

Eighty-one (student) teachers took part in the study. However, data from six participants were removed from the analysis because they did not consent to further data processing. In total, $N = 75$ participants of whom $n = 53$ were in-service teachers and $n = 22$ were student teachers from German universities and schools agreed to allow their data to be processed and participate in the study. The mean age of the participants was $M = 35.87$ ($SD = 9.13$, Range: 20 to 58). The survey design of the study was cross-sectional, and it was conducted through an online survey through UniPark in the winter of 2019 (https://www.unipark.com). The sample included $n = 43$ (57%) female and $n = 32$ (43%) male participants. The in-service teachers predominantly taught in lower track secondary schools (Realschule and Mittelschule, $n = 26$), followed by grammar schools (Gymnasium, $n = 16$), vocational schools ($n = 4$), schools for children with special needs ($n = 3$), elementary schools ($n = 2$), and other schools ($n = 2$). The in-service teachers had an average of 14.53 years of teaching experience (Range: 2 to 44 years). On average, student teachers were in their 7th and 5th semesters of study (Range: 1 to 16 semesters, mode = 13). Participants were invited by email, through distribution lists from their respective networks, or social media posts. After participants gave their consent for data processing, they were first instructed to answer demographic questions. Then, the online survey continued with the self-assessment questionnaire, which contained scenario-based vignettes, followed by the objective assessment of TK. On average, participants completed the objective assessment and scenario-based self-assessment test in 25 min. Table 1 presents descriptive statistics for the scores related to the objective assessment and the scenario-based self-assessment of TK. On average, (student) teachers scored $M = 117.6$ ($SD = 8.64$) out of 136 on the objective assessment test of TK. Their scenario-based self-assessments regarding the five subcomponents ranged on average from $M = 1.08$ (SD = .73, producing and presenting information using technology) to $M = 2.50$ (SD = .59, Searching for and processing information using technology) on the five-point Likert scale.

Measures

Scenario-based self-assessment of TK

Scenario-based self-assessment—one for each of the five subcomponents—were used for the self-assessment of TK (for detailed information about scenario-based self-assessment, see Sailer et al., 2021). The scenarios included everyday situations that depicted technology use and were based on the five subcomponents. For each scenario, participants were asked to indicate whether they had the knowledge and skills to react adequately to the described situation and, in addition, whether they were able to advise others regarding the situation. The subcomponents were assessed with a total of 60 items, 12 for each of the five subcomponents. Items were rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The five subcomponents included (a) operating and applying technology (Cronbach's α = .93), (b) searching for and processing information with technology (α = .93), (c) communicating and collaborating with technology (α = .86), (d) producing and presenting information with technology (α =. 94), and (e) analyzing and reflecting information with technology (α = .94). Tables 2 and 3 present sample

Table 1. Descriptive statistics for objective assessments and scenario-based self-assessments of TK.

	Min.	*Max.*	*M*	*SD*
Technological knowledge (TK)—objective assessment	94.0	130.0	117.6	8.64
Technological knowledge (TK)—self-assessment				
Operating and applying technology	2.17	5.00	4.25	.69
Searching for and processing information with technology	2.50	5.00	4.35	.59
Communicating and collaborating with technology	1.67	5.00	4.18	.83
Producing and presenting information with technology	1.08	5.00	4.18	.73
Analyzing and reflecting information with technology	1.92	5.00	3.91	.79

Table 2. Scenario-based assessment item—operating and applying technology.

You are planning a long train journey, and you want to provide yourself with reading material for the ride. To reduce your luggage, you borrow a friend's tablet to read eBooks on it. However, you have to familiarize yourself with the operation and application of the tablet first, especially pertaining to the circumstances involved in a train ride.					
Based on the scenario described above, please rate the following statements on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*)					
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
I have the knowledge to successfully use applications on technical devices	1	2	3	4	5
I can successfully use applications on technical devices	1	2	3	4	5
I am able to support others successfully to use applications on technical devices	1	2	3	4	5

Note. The item was translated from German.

Table 3. Scenario-based self-assessment item—producing and presenting information with technology.

You want to contribute to the public relations work in your place of employment, so you decide to create an informational video about your work. Your place of employment has diverse, appropriate hardware and software for creating and editing videos.					
Based on the scenario described above, please rate the following statements on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*)					
	Strongly disagree	Disagree	Neither agree or disagree	Agree	Strongly agree
I have the knowledge to create engaging and effective media products	1	2	3	4	5
I can create engaging and effective media products	1	2	3	4	5
I am able to help others create engaging and effective media products	1	2	3	4	5

Note. The item was translated from German.

items for the subcomponents *operating and applying technology* and *producing and presenting information with technology*.

Objective assessment of TK

In order to comprehensively measure TK of teachers, the constructed test for objective measurement of TK contains questions on both theoretical and applied knowledge of TK of teachers derived from the KMK framework (2019). According to the KMK (2019b), teachers should be able to name basic theoretical components of computers and provide information about how computer systems, networks, data, and standard software work. The areas of theoretical knowledge refer to (1) knowledge about the structure and components of a computer, (2) knowledge about software, (3) knowledge about networks, and (4) knowledge in the area of data. In addition, according to the KMK specifications (2019), teachers should also be able to use TK to apply their knowledge in an application-oriented manner in the classroom, e.g., to identify sources of error when using technologies or to use software as needed. Accordingly, the test for the objective measurement of TK contains questions on (1) the use of computer systems, (2) the use of software, and (3) the use of networks. To ensure the quality and the validity of the objective assessment test, researchers ($N = 2$, post-Doc level, $N = 1$ senior PhD student) helped to develop and continuously improve the objective assessment of TK by providing their expert feedback. All items were considered suitable to measure valid TK of (prospective) teachers in feedback loops in the context of expert ratings. The objective test on TK consisted of 34 multiple-choice tasks with four answer options, of which one to four could be correct. Points could also be earned by not checking wrong answers. A total of 136 points were possible. Two sample questions from the objective assessment test are presented in Tables 4 and 5. Based on

Table 4. Example item from the objective test of TK.

Tick any of the following statements that are true
a) An HDMI cable can be used to connect to the Internet between the router and the computer. b) *An HDMI cable is used to transmit audio and video signals.* c) *A USB key is a type of external data storage.* d) *A USB cable can be used to connect computers to external devices.*

Note. The item was translated from German. Correct answers are marked in italics.

Table 5. Example item from the objective test of TK.

You want to do research on the Internet. How do you proceed?
a) I open a browser window and log in to my email account. Then I enter the search term in the search bar. b) I open the search function on my computer and enter the search term. A browser window opens with a list of research results. c) *I open a browser window, enter a search engine website, and enter what I want to search for with the search engine.* d) I open my Internet app and then enter search terms in any application.

Note. The item was translated from German. The correct answer is marked in italics.

previous research instruments for the objective assessment of TK (e.g., Test of Technological and information Literacy, TILT; see Senkbeil et al., 2013), the objective assessment test was designed to be unidimensional and to include aspects of knowledge and skills related to TK. To test the one-dimensionality of the objective TK measure, we used the Bayesian Information Criterion (BIC) and the Akaike's Information Criterion (AIC) as indicators of model fit (Rost, 2004). Lower indices indicate better model fit. We compared the one-dimensional model (BIC: 4075; AIC: 3904) with the two-dimensional model (BIC: 4226; AIC: 3901) and the three-dimensional model (BIC: 4313; AIC: 3915). The BIC was lowest for the one-dimensional model. Although the two-dimensional model had a slightly lower AIC, the literature recommends choosing the model with the lower dimension if the difference is less than 10% (Rost, 2004). We therefore conclude, based on the model fit indices, that the one-dimensionality of the objective TK measure is supported. We used Rasch analysis to assess person-ability scores for the objective TK measure. The cutoff values (0.5–1.5) proposed by Linacre (2002) for Infit-MNSQ and Outfit-MNSO 0.5–1.5 are supported. Rasch analysis of the unidimensional model yielded good estimates of item response reliability (WLE) (.73).

Statistical analysis

Path analysis was used to test the hypothesized structural relationships between the person-ability score of the objectively assessed and scenario-based self-assessed TK of both in-service teachers and student teachers (using the lavaan package; Rosseel, 2012). The goodness-of-fit measures for the path analysis were based on Hu and Bentler (1999) recommendations: Root Mean Square Error of Approximation (RMSEA) < .06, Standardized Root Mean Square Residual (SRMR) < .08, Comparative Fit Index (CFI) > .90, and Chi-Square (χ^2) > .05. We conducted all analyses in R version 4.0.5.

Results

A path analytic model was tested to examine the relationships between the scenario-based self-assessment measures and the objective measure of TK. The model fit indices showed a fully saturated model, $\chi^2(5) = 22.39$, CFI > .99, SRMR < .01, RMSEA < .01, although the p-value of the chi-square test was significant. Heene et al. (2011) showed that unique variances affect the value of the chi-square test of the model when the sample size is small. Therefore, we concluded that the model tested in this study was acceptable. The results showed that scenario-based self-assessment measured with the subcomponent operating and applying technology ($\beta = 0.25$, SE = 0.13, p =.05) significantly predicted the objective assessment of TK (R^2

= .23). However, the analysis showed that scenario-based self-assessment measured with the subcomponents producing and presenting information with technology ($\beta = 0.21$, SE = 0.12, p =.08), searching for and processing information with technology (β = −0.05, SE = 0.14, p = .72), communicating and collaborating (β = −0.10, SE = 0.18, p = .58), and analyzing and reflecting on information with technology ($\beta = 0.07$, SE = .11, p = .52) did not significantly predict teachers' objective assessment of TK. All standardized regression coefficients are shown in Figure 1. Table 6 presents correlation coefficients for the five subcomponents of scenario-based self-assessment in TK. All of the six scales (objective assessment of TK and self-assessment of TK) significantly correlated with each other but differed substantially in their magnitude with the range of r = .28 to $r = 73$. Overall, the results from the path analysis supported the assumption that scenario-based self-assessment items concerning the functional use of technology predicted teachers' objectively assessed TK.

Discussion

Previous work has regularly shown that there is no significant relationship between self-assessments and objective assessment measures of teachers' technology use (Akyuz, 2018, Baier & Kunter, 2020; Drummond & Sweeney, 2017), suggesting that self-assessments may be biased and inaccurate. Based on previous research, scenario-based self-assessment is a promising scaffold to approximate the results of objective assessment measures related to personal behavior, as it is easier for

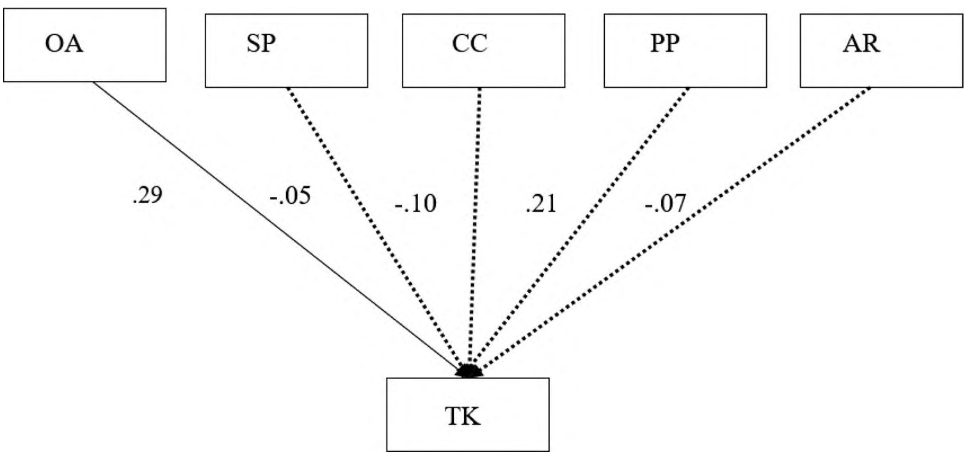

Figure 1. Estimated model in which the subcomponents of the scenario-based self-assessment of TK were used to predict the objective assessment of TK for in-service and student teachers.*Note*. TK = the objective assessment of technological knowledge. The predictors were self-assessed: OA = operating and applying technology; SP = searching for and processing information with technology; CC = communicating and collaborating with technology; PP = producing and presenting information with technology; AR = analyzing and reflecting on information with technology. Nonsignificant paths are represented with dotted lines.

Table 6. Estimated correlation matrix of the five subscales regarding the scenario-based self-assessment and objective assessment of TK.

	1	2	3	4	5	6
1. Technological knowledge (TK)—objective assessment	–					
Technological knowledge (TK)—self-assessment						
2. Operating and applying technology	.46**	–				
3. Searching for and processing information with technology	.26*	.54**	–			
4. Communicating and collaborating with technology	.35*	.66**	.64**	–		
5. Producing and presenting information with technology	.46*	.69**	.52**	.73**	–	
6. Analyzing and reflecting on information with technology	.36*	.56**	.58**	.69**	.66**	–

* $p < .05$.
** $p < .01$.

subjects to assess their own behavior in a specific situation than to accurately represent their behavior using self-assessment scales, as their responses may depend on their subjective reality (Van Soest et al., 2011). This assumption has also been postulated by scholars in educational science (Scheiter, 2021; Talsma et al., 2018), as scenario-based self-assessment could be a way to better assess one's own skills and, accordingly, avoid bias when comparing self-assessment with actual objectively measured skills. Consequently, this study examined how the results of scenario-based self-assessments aligned with the results of objective assessment measures related to teachers TK.

We found that one of the five core components related to teachers' digital skills –*operating and applying technology*—used to measure the scenario-based self-assessment of TK explained a substantial share of the variance (23%) in the TK measure. In addition, we found a tendency for the subcomponent *producing and presenting information with technology* (p = .08), although the results were not significant. According to KMK (2019a, 2019b), the components *operating and applying technology* and *producing and presenting Information with technology* are closely related to operational technical use of technology. According to Fraillon et al. (2020), operational technical use of technology is one of the less demanding levels of skills related to technology use. However, internationally established theoretical frameworks and meta-analyses (Ferrari, 2013; Redecker, 2017; Siddiq et al., 2016) have highlighted the relevance of technical operational skills as a prerequisite for successful technology use.

Although the results should be interpreted with caution due to the explorative nature of the study, the significant relationship between the scenario-based self-assessment of the subcomponent *operating and applying technology* and the objective TK measure suggests that scenario-based self-assessment might be an appropriate tool to approach the results objective measures, especially for technical operational skills. Accordingly, the tendency of the subcomponent *Producing and Presenting Information with technology* could also be interpreted toward the same direction, because producing and presenting information with technology makes up a large part of skills regarding the operational technology use (KMK, 2019b).

Because previous research has shown non substantial relationships between self-assessment and objective assessment in educational research (Parry et al., 2021), scholars have often called for increased use of objective assessment instruments (e.g., Lachner et al., 2019) to represent participants' actual abilities. However, there are major criticisms of objective assessments of TK: First, objective assessments of TK tend to be very specific (e.g., about how a particular program works; Petko, 2020). Second, objective assessment tests become outdated very quickly due to technological advances, including those related to teaching and learning processes (Siddiq et al., 2016). Moreover, previous research has already shown that especially more complex skills (e.g., evaluating information using information technologies; Fraillon et al., 2020) are less suitable for measurement with objective measures because objective assessment measures place high demands on test takers' knowledge and skills, often leading to reluctance and dropout compared with self-assessment measures (Kleinert et al., 2015). Furthermore, Siddiq et al. (2016) criticized objective assessment for being limited to restricted response formats (e.g., multiple-choice), which limits the full potential of objective assessment measures. The full potential of objective assessments for digital skills will not be realized until authentic simulations can be created on the computer to assess, for example, the more complex digital skills that are required to successfully participate in an information-based society (Fraillon et al., 2020). However, authentic situations have rarely been implemented in research due to their complexity. Therefore, real-life situations are usually mimicked (e.g., by using screenshots to measure digital skills; Siddiq et al., 2016). Consequently, the objective measurement of digital skills seems to be a challenge for research. Current approaches for objectively measuring teachers' digital skills include analyzing (student) teachers' lesson plans (e.g., Backfisch et al., 2020). Lesson plan analysis is a valuable measure for providing a proxy for the quality of technology integration (Backfisch et al., 2020). Nonetheless, lesson plans are an indirect measure that, according to Petko (2020), leaves a great deal of room for heterogeneous interpretations, even though researchers have reported high interrater reliability.

Because previous research has shown a weak relationship between self-assessment and objective assessment in educational research (Parry et al., 2021), scholars have often called for increased use of objective assessment instruments (e.g., Lachner et al., 2019) to represent participants' actual abilities. However, there are major criticisms of objective assessments of TK: First, objective assessments of TK tend to be very specific (e.g., about how a particular program works; Petko, 2020). Second, objective assessment tests become outdated very quickly due to technological advances, including those related to teaching and learning processes (Siddiq et al., 2016). Moreover, previous research has already shown that especially more complex skills (e.g., evaluating information using information technologies; Fraillon et al., 2020) are less suitable for measurement with objective measures because objective assessment measures place high demands on test takers' knowledge and skills, often leading to reluctance and dropout compared with self-assessment measures (Kleinert et al., 2015). Furthermore, Siddiq et al. (2016) criticized objective assessment for being limited to restricted response formats (e.g., multiple-choice), which limits the full potential of objective assessment measures. The full potential of objective assessments for digital skills will not be realized until authentic simulations can be created on the computer to assess, for example, the more complex digital skills that are required to successfully participate in an information-based society (Fraillon et al., 2020). However, authentic situations have rarely been implemented in research due to their complexity. Therefore, real-life situations are usually mimicked (e.g., by using screenshots to measure digital skills; Siddiq et al., 2016). Consequently, the objective measurement of digital skills seems to be a challenge for research. Current approaches for objectively measuring teachers' digital skills include analyzing (student) teachers' lesson plans (e.g., Backfisch et al., 2020). Lesson plan analysis is a valuable measure for providing a proxy for the quality of technology integration (Backfisch et al., 2020). Nonetheless, lesson plans are an indirect measure that, according to Petko (2020), leaves a great deal of room for heterogeneous interpretations, despite the fact that researchers have reported high interrater reliability.

Although more objective assessment measures are often called for in the current literature, they are difficult to implement and often lead to reluctance. However, according to the results of this study, scenario-based self-assessment with a description of a concrete situation seems promising for more closely approximating the results of objective assessment measures. Furthermore, according to Sailer et al. (2021), scenario-based self-assessment has two advantages over regular self-assessment: First, a "common standard" (p. 4) can be achieved for respondents via a specific scenario, so that, for example, social desirability bias can be reduced. Second, vignettes in a scenario-based self-assessment can be used to assess multiple aspects (e.g., situations where relevant knowledge and skill domains need to be applied) rather than only asking the participant to assess their own knowledge or skills.

Limitations and future directions

Due to the small number of participants and the predominance of in-service teachers in the sample (71%), it was not possible to independently compare objectively assessed TK and the scenario-based self-assessment of TK for in-service teachers and student teachers. Future work should therefore include personal and contextual factors (e.g., see Aesaert et al., 2017; Lucas et al., 2021; Ober et al., 2021; Schmid et al., 2021) to determine whether the findings of this study hold, as this was not possible to do in the current study due to missing values and the small sample size. Second, to draw concrete conclusions about the effectiveness of scenario-based self-assessments, follow-up work could contrast both scenario-based assessment measures and regular self-assessment measures with objective assessment measures. Moreover, longitudinal studies could be conducted to provide detailed insights into the extent to which scenario-based self-assessment measures and objective assessment measures are consistent. Although the objective assessment measure of TK demonstrated good psychometric quality, the objective assessment measure was developed for the purpose of the study to capture the unidimensional functional use of technology and therefore needs to be developed further in future studies. Future studies could also use objective assessment measures that cover multiple dimensions of technology use and compare these with scenario-based self-assessment measures.

Conclusion

Self-assessment has often been criticized in educational research for failing to measure individuals' actual abilities and for being influenced by factors such as social desirability (Hatlevik et al., 2018), resulting in current calls for the use of objective assessments in addition to self-assessment measures (Lachner et al., 2019; Seufert et al., 2021). Nonetheless, objective assessments are difficult to use in practice because participants are often reluctant to complete them, often resulting in high dropout rates (Kleinert et al., 2015). According to our results, scenario-based self-assessment seems to be a promising approach for providing results that are close to those from objective assessment measures regarding technical operational skills.

Notes on Contributors

Tamara Kastorff is Research Associate and doctoral candidate at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Michael Sailer is Intermediate Professor for digital teaching and learning processes at the FAU University and senior lecturer at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Johanna Vejvoda is Research Associate at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Florian Schultz-Pernice is a postdoctoral scholar and senior lecturer at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Valentin Hartmann is a master candidate in learning sciences and assistant at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Anna Hertl is an in-service teaching professional and graduate at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Sonja Berger is Research Associate and doctoral candidate at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Karsten Stegmann is a postdoctoral scholar and senior lecturer at the Department of Psychology, Ludwig-Maximilians-University of Munich.

Funding

This work was supported by German Federal Ministry of Research and Education.

ORCID

Tamara Kastorff http://orcid.org/0000-0002-7603-304X
Michael Sailer http://orcid.org/0000-0001-6831-5429

References

Aesaert, K., Voogt, J., Kuiper, E., & van Braak, J. (2017). Accuracy and bias of ICT self-efficacy: An empirical study into students' over- and underestimation of their ICT competences. *Computers in Human Behavior*, *75*(1), 92–102. https://doi.org/10.1016/j.chb.2017.05.010

Akyuz, D. (2018). Measuring technological pedagogical content knowledge (TPACK) through performance assessment. *Computers & Education*, *125*(4), 212–225. https://doi.org/10.1016/j.compedu.2018.06.012

Backfisch, I., Lachner, A., Hische, C., Loose, F., & Scheiter, K. (2020). Professional knowledge or motivation? Investigating the role of teachers' expertise on the quality of technology-enhanced lesson plans. *Learning and Instruction*, *66*(2), 101300. https://doi.org/10.1016/j.learninstruc.2019.101300

Baier, F., & Kunter, M. (2020). Construction and validation of a test to assess (pre-service) teachers' technological pedagogical knowledge (TPK). *Studies in Educational Evaluation*, *67*, 100936. https://doi.org/10.1016/j.stueduc.2020.100936

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Drummond, A., & Sweeney, T. (2017). Can an objective measure of technological pedagogical content knowledge (TPACK) supplement existing TPACK measures? *British Journal of Educational Technology*, *48*(4), 928–939. https://doi.org/10.1111/bjet.12473

Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106. https://doi.org/10.1111/j.1529-1006.2004.00018.x

Ferrari, A. (2013). DIGCOMP: A framework for developing and understanding digital competence in Europe. In Y. Punie & B. N. Brecko (Eds.), *JRC scientific and policy reports* (pp. 4–32). Publications Office of the European Union. https://doi.org/10.2788/52966.

Forschungsgruppe Lehrerbildung Digitaler Campus Bayern [Research Group Teacher Education Digital Campus Bavaria]. (2017). Kernkompetenzen von Lehrkraeften für das Unterrichten in einer digitalisierten Welt [Core competencies of teachers for teaching in a digital world]. *Merz Medien + Erziehung: Zeitschrift für Medienpadagogik*, *4*, 65–74.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 International Report* (1st ed.) Springer International Publishing. Imprint: Springer. https://doi.org/10.1007/978-3-030-38781-5

Hatlevik, I. K. R., & Hatlevik, O. E. (2018). Examining the relationship between teachers' ICT self-efficacy for educational purposes, collegial collaboration, lack of facilitation and the use of ICT in teaching practice. *Frontiers in Psychology*, *9*, 935. https://doi.org/10.3389/fpsyg.2018.00935

Hatlevik, O. E., Throndsen, I., Loi, M., & Gudmundsdottir, G. B. (2018). Students' ICT self-efficacy and computer and information literacy: Determinants and relationships. *Computers & Education*, *118*(1), 107–119. https://doi.org/10.1016/j.compedu.2017.11.011

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16*(3), 319–336. https://doi.org/10.1037/a0024917

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, *98*(1), 191–207. https://doi.org/10.1007/978-3-531-91826-6_16 https://doi.org/10.1017/S000305540400108X

King, G., & Wand, J. (2007). Comparing incomparable survey responses: Evaluating and selecting anchoring vignettes. *Political Analysis*, *15*(1), 46–66. https://doi.org/10.1093/pan/mpl011

Kleinert, C., Christoph, B., & Ruland, M. (2015). Auswirkungen der Administration von Kompetenztests im Rahmen einer Panelerhebung für Erwachsene. [Effects of administering competency tests in an adult panel survey]. In J. Schupp & C. Wolf (Eds.), *Nonresponse bias* (pp. 359–381). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-10459-7_11

Kultusministerkonferenz (KMK) [conference of federal ministers]. (2019a). *Bildung in der digitalen Welt. Strategie der Kultusministerkonferenz [Education in a digital world. Strategy of the conference of federal ministers of Germany]*. https://www.kmk.org/themen/bildung-in-der-digitalen-welt/strategie-bildung-in-der-digitalen-welt.html

Kultusministerkonferenz (KMK) [conference of federal ministers]. (2019b). *The Education System in the Federal Republic of Germany 2017/2018 A description of the responsibilities, structures and developments in education policy for the exchange of information in Europe*. https://www.kmk.org/fileadmin/Dateien/pdf/Eurydice/Bildungswesen-engl-pdfs/dossier_en_ebook.pdf

Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70. https://www.researchgate.net/publication/241616400_What_Is_Technological_Pedagogical_Content_Knowledge#fullTextFileContent

Lachner, A., Backfisch, I., & Stürmer, K. (2019). A test-based approach of modeling and measuring technological pedagogical knowledge. *Computers & Education*, *142*(1), 103645. https://doi.org/10.1016/j.compedu.2019.103645

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85–106.

Lucas, M., Bem-Haja, P., Siddiq, F., Moreira, A., & Redecker, C. (2021). The relation between in-service teachers' digital competence and personal and contextual factors: What matters most? *Computers & Education*, *160*(3), 104052. https://doi.org/10.1016/j.compedu.2020.104052

Mourlam, D., Chesnut, S., & Bleecker, H. (2021). Exploring preservice teacher self-reported and enacted TPACK after participating in a learning activity types short course. *Australasian Journal of Educational Technology*, *37*(3), 152–169. https://doi.org/10.14742/ajet.6310

Ober, T. M., Hong, M. R., Rebouças-Ju, D. A., Carter, M. F., Liu, C., & Cheng, Y. (2021). Linking self-report and process data to performance as measured by different assessment types. *Computers & Education*, *167*, 104188. https://doi.org/10.1016/j.compedu.2021.104188

Parry, D. A., Davidson, B. I., Sewall, C. J. R., Fisher, J. T., Mieczkowski, H., & Quintana, D. S. (2021). A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour*, *5*(11), 1535–1547. https://doi.org/10.1038/s41562-021-01117-5

Petko, D. (2020). Quo vadis TPACK? Scouting the road ahead. In: Bastiaens, Theo. *EdMedia + Innovate Learning* (pp. 1349–1358). Association for the Advancement of Computing in Education (AACE).

Peura, P., Aro, T., Viholainen, H., Räikkönen, E., Usher, E. L., Sorvo, R., & Aro, M. (2019). Reading self-efficacy and reading fluency development among primary school children: Does specificity of self-efficacy matter? *Learning and Individual Differences*, *73*, 67–78. https://doi.org/10.1016/j.lindif.2019.05.007

Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu* (NoJRC107466). Joint Research Centre (Seville site).

Rohatgi, A., Scherer, R., & Hatlevik, O. E. (2016). The role of ICT self-efficacy for students' ICT use and their achievement in a computer and information literacy test. *Computers & Education*, *102*(4), 103–116. https://doi.org/10.1016/j.compedu.2016.08.001

Rosseel, 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

Rost, J. (2004). *Testtheorie—Testkonstruktion [Test theory—Test construction]*. Hans Huber.

Rubach, C., & Lazarides, R. (2021). Addressing 21st-century digital skills in schools—Development and validation of an instrument to measure teachers' basic ICT competence beliefs. *Computers in Human Behavior*, *118*, 106636. https://doi.org/10.1016/j.chb.2020.106636

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*(2), 106625. https://doi.org/10.1016/j.chb.2020.106625

Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung [Technology-enhanced learning and teaching: An overview]. *Zeitschrift fur Erziehungswissenschaft*, *24*, 1–22. https://doi.org/10.1007/s11618-021-01047y

Scherer, R., & Siddiq, F. (2015). Revisiting teachers' computer self-efficacy: A differentiated view on gender differences. *Computers in Human Behavior*, *53*, 48–57. https://doi.org/10.1016/j.chb.2015.06.038

Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education*, *88*, 202–214. https://doi.org/10.1016/j.compedu.2015.05.005

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*(4), 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 (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, *157*(4), 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*(4), 106586. https://doi.org/10.1016/j.chb.2020.106586

Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013). The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*, *5*, 139–161.

Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, *115*, 106552. https://doi.org/10.1016/j.chb.2020.106552

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14. https://doi.org/10.3102/0013189X015002004

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past—A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, *19*(4), 58–84. https://doi.org/10.1016/j.edurev.2016.05.002

Talsma, K., Schüz, B., Schwarzer, R., & Norris, K. (2018). I believe, therefore I achieve (and vice versa): A meta-analytic cross-lagged panel analysis of self-efficacy and academic performance. *Learning and Individual Differences*, *61*, 136–150. https://doi.org/10.1016/j.lindif.2017.11.015

van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjctive questions. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)*, *174*(3), 575–595. https://doi.org/10.1111/j.1467-985X.2011.00694.x

van Vliet, P. J., Kletke, M. G., & Chakraborty, G. (1994). The measurement of computer literacy: A comparison of self-appraisal and objective tests. *International Journal of Human-Computer Studies*, *40*(5), 835–857. https://doi.org/10.1006/ijhc.1994.1040