International Journal of Technology and Design Education (2023) 33:1265–1284
https://doi.org/10.1007/s10798-022-09774-x

Fostering pre-service primary school teachers' ability to recognize differences in pupils' understanding of technical systems

Dannie Wammes[1] · Bert Slof[2] · Willemijn Schot[3] · Liesbeth Kester[4]

Accepted: 1 August 2022 / Published online: 24 August 2022

Abstract
Pupils benefit from adaptive instruction and feedback from their teachers. A prerequisite for providing adaptive instruction is that teachers' diagnostic ability enables them to correctly perceive their pupils' skill level. A short course has been developed to improve primary school teachers' diagnostic ability for engineering. Based on Nickerson's anchoring and adjustment model, the participants became aware of the differences their own and pupils' use of information when constructing technical systems. The Fischer scale was used as a model to understand and identify pupils' development in using such information. The participants were given examples of pupils' reconstructions of technical systems. They were asked to evaluate these work products in four ways: relative and absolute, combined with intuitive and explicit. The results reveal that relative and absolute diagnoses can differ considerably for the same teacher and between teachers, depending on whether they are implicit or explicit. Post-test results show that the course improved the ability to explain the differences between pupils' use of information to construct a technical system. The course also had a strong, significant, positive impact on teachers' self-efficacy beliefs about technology education.

Keywords Primary education · Diagnostic ability · Technical systems · Training

✉ Dannie Wammes
dannie.wammes@han.nl

[1] HAN University of Applied Sciences, HAN PABO, Kapittelweg 35, 6525 EN Nijmegen, The Netherlands

[2] National Institute Curriculum Development, SLO, Enschede, The Netherlands

[3] Educational Consultancy and Professional Development, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands

[4] Department of Education, Faculty of Social and Behavioural Sciences, Utrecht University, Utrecht, The Netherlands

Introduction

Technology education is part of the primary education curriculum in many countries, whether or not integrated into STEM. Learning outcomes in this domain can be improved when teachers align their instruction and feedback with pupils' prior knowledge (Behrmann & Souvignier, 2013; Hattie, 2013). Such alignment requires a correct diagnosis of pupils' ability level (Shavelson, 1978; Black & Wiliam, 1998; Van de Pol et al., 2010). Most primary school teachers, and even those with considerable experience in teaching technology, lack sufficient insight into the technical abilities of their pupils. As a result, they doubt the quality of their support for the optimal development of these abilities (Moreland & Jones, 2000; Scharten & Kat-de Jong, 2012). This study examines what a short course for prospective teachers can contribute to their diagnostic ability and whether this would impact their technology education self-efficacy. This is important since correct diagnoses of pupils' proficiency levels are the key to effectively adapting instruction, tasks and feedback to differences between pupils.

Diagnostic ability

In this study, we consider the diagnostic ability of teachers as a combination of their judgement accuracy and ability to explain and communicate their diagnoses. Teacher judgements can be relative or absolute (Südkamp et al., 2012). Ranking pupils by their results is a relative type of judgement. The accuracy of a relative judgement is usually expressed as the correlation between the teachers' estimate and pupils' rank based on objective criteria. An absolute judgement is normative, for example, a teacher's estimate of the position of a pupil on a developmental scale or their test performance (Schrader & Helmke, 2001). The accuracy of absolute judgements can be expressed in different ways. Still, like relative judgement accuracy, it is usually expressed as a correlation here between the teachers' estimate and the pupils' results. Südkamp et al. concluded that teachers tend to be more accurate in their relative than in their absolute diagnoses. Only a weak relationship has been found between these types of judgement (Dunlosky & Thiede, 2013).

In addition to relative or absolute, a diagnosis can be implicit, based on intuition, or explicit, based on consciously and communicably weighing up the information (Wood, 2014). Wood argues that an implicit, intuitive judgement, which often arises from a first impression, could be based on the brain's fast, automatic System 1 processes (Kahneman, 2011), whereas an explicit evaluation will be primarily based on System 2 processes. Differences in accuracy could, according to Wood, relate to the interaction between the two processes. For example, there is evidence that a first impression (System 1) influences the choice of questions (System 2) asked during an assessment (Govaerts et al., 2013).

Pupils benefit most from teachers whose diagnoses are accurate and explicit (Edelenbos & Kubanek-German, 2004). Such diagnoses enable effective differentiated instruction and feedback (Van de Pol et al., 2010). The ability to diagnose correctly and explicitly requires knowledge about how a particular skill develops and how that can be recognised in pupils' activities (Gingerich et al., 2014; Jones & Moreland, 2004).

Assessing pupils' technical skills is a complex endeavour for primary school teachers. Even within the subdomain of engineering, which is explored in this study, a wide variety of activities and associated skills exist (Pearson & Young, 2002). One of the overarching characteristics of engineering is that most of these activities relate to systems,

e.g., constructions, pneumatic, mechanical and electrical systems, and ICT. In primary technology education, many activities are about such systems, ranging from building walls and roads to robotics with Lego Mindstorm (Brophy et al., 2008; Mullis & Martin, 2017; National Assessment Governing Board, 2013; Svensson et al., 2012).

Pupils' understanding of these systems develops hierarchically, from identifying the components of a system to the ability to imagine the systems' behaviour over time (Assaraf & Orion, 2010). This hierarchy is based on an increasing ability to combine knowledge about the system's components, interactions and functions. A similar hierarchy typifies Fisher's developmental model (1980). Therefore, Sweeney and Sterman (2007) propose to use Fischers' model to interpret pupils' development in their understanding of technical systems.

Teacher characteristics and diagnostic ability

Teachers in primary education find it difficult to infer pupils' level of understanding of technical systems (Wammes et al., 2022). In their technology lessons, they tend to focus their feedback on other topics, like pupils' ability to cooperate or their mathematical skills (Moreland & Jones, 2000). Assumed causes include limited knowledge of technology (Jones & Moreland, 2004; Rohaan et al., 2012; Sanjosé & Otero, 2021) and insufficient knowledge about developing complex thinking skills (Retnawati et al., 2018), both of which are needed to understand technical systems. Other teacher characteristics related to diagnostic ability are self-efficacy beliefs (De Paulo et al., 1997), work experience (Ready & Wright, 2011; Wammes et al., 2022) and intelligence (Kaiser et al., 2012). These other characteristics only seem to explain teachers' diagnostic ability to a limited extent (see review Urhahne & Wijnia, 2021). It can be assumed that self-efficacy beliefs do not improve diagnostic ability; rather, they are affected by it.

Thus, opportunities to enhance teachers' diagnostic abilities in the context of technical systems lie primarily in broadening teachers' technical knowledge and their knowledge about how pupils develop their understanding of systems. Additionally, it can be expected that a course will be more effective when the content is linked to the participants' interests (Nauta et al., 2002). Generally, primary school teachers are particularly interested in developing their pupils' skills (Butler, 2012) and less in technical knowledge (Hsu et al., 2011; Knezek et al., 2011). Therefore, a course that aims to improve primary school teachers' diagnostic ability in the field of technology should make pupils' learning the focal point and introduce technical knowledge within that context.

Course design

Time for courses for teachers in primary education is scarce, but as Ostermann et al. (2018) have demonstrated, even short courses can positively affect diagnostic ability. Ostermann et al. designed their course about diagnosing pupils' ability to interpret graphs using Nickerson's (1999) anchoring and adjustment model. This model describes how we construct our ideas about what others know. It predicts that we tend to think that others think like ourselves. Ostermann et al. used Nickerson's model to improve teachers' diagnostic abilities in three steps. First, they made the participating teachers aware of their strategies for interpreting graphs. Then, they showed them that their knowledge was incomparable to their pupils' knowledge. Finally, they created awareness of the task responses commonly shown by the pupils in their classes.

The course in this study followed the same steps as Ostermann et al. in three meetings of one and a half hours, including about 50 min for pretest and posttest. First, the teachers were made aware of their understanding of how technical systems function by asking them to construct an electrical and a mechanical system while thinking aloud. How their thinking was incomparable to that of their pupils in primary classrooms was demonstrated by showing them the results of pupils on the same tasks. Finally, common responses from pupils in primary school were shown, categorised and explained by the different phases of Fischers' skill development scale (Fischer, 1980; Van der Steen, 2014; Wammes et al., 2021).

The Fischer scale consists of *three main phases*. The first phase (sensorimotor) is characterised by actions solely based on sensorimotor information. The second phase (representation) evolves out of repeated experiences and their neurological effects, which create the ability to remember what happened in previous situations and to include that knowledge in a choice of action (Edelman, 1992; Thelen & Smith, 1994). The third phase (abstraction) stems from the successive and repeated combination of multiple representations. In this phase, pupils are able, for a specific phenomenon, to identify its main characteristics and use these to choose an appropriate action in situations that have not been previously encountered. The first signs of actions or utterances based on reasoning at such an abstract level are usually seen between 12 and 14 (Fischer & Bidell, 2007; Molnár et al., 2017; Fischer, 1980).

Within each phase, Fischer distinguishes *three recursive levels*. The first level is a single piece of sensorimotor information, a single representation or a single abstraction that directs a pupil's action. The second level is known as 'mapping', which indicates the combined use of sensorimotor information in the first phase and the combined use of representations in the second phase. As the combined use of abstractions is very uncommon among pupils in primary education, the explanation of the Fischer scale was restricted to the level of a single abstraction. The third recursive level is called 'system' and indicates multiple combined sets of sensorimotor information or representations.

The Fischer scale was introduced in the course using examples of verbal utterances of learners described in several studies featuring the Fischer scale (Bassano & Van Geert, 2007; Meindertsma et al., 2014; Van der Steen, 2014) and then applied to pupils' attempts to restore an electrical and a mechanical system. The examples used provided an overview of the development of the use of information to reconstruct both systems.

Particular attention was paid to affordances (Chemero, 2003; Gibson, 1977). Affordances are found in the relationship between the properties of an object or situation and the possibility of an organism perceiving them. Affordances impose a strong, often unconscious, influence on a choice of action. An example of this is the graphical objects on a computer screen. A button shape usually results in pressing, while a bar will elicit scrolling. Programming a button to react to scrolling would cause a lot of confusion.

Affordances play an important role in the way pupils interact with technical systems. They may result in effective actions but can trigger ineffective actions. Affordances play a major role in the first phase of the Fischer scale because, in this phase, actions are based on the available information. In the second phase, affordances are increasingly weighted by their possible function in the entire system (Svensson et al., 2012; Sweeney & Sterman, 2007). For instance, in constructing an electric circuit, most pupils tend to put clips on the outer points of connectors, even when these are insulated and even when these pupils correctly answer a multiple-choice question about the effect of conducting and insulating materials in an electric circuit. Through learning and experiences, pupils will gradually ignore their inclination to connect to outer points and only consider metal connections

appropriate. During the course, participants learned to recognise the role of affordances within this developmental process.

When diagnosed on a certain level of the Fischer scale, what is needed to bring pupils' thinking forward was discussed in the course's third session. Making pupils aware of the unconscious role of affordances in decisions on actions is especially useful for pupils who are inclined to react without reflecting on the systems' function. Providing more experience and emphasising past experiences is especially needed for those with work products at the sensorimotor level. Encouragement and help to explain what guided the construction of work products are important for pupils who can construct work products at a high level. Such work products are often intuitively constructed, and discussing them helps pupils develop the language that can bring their understanding of technical systems forward.

Research question

This article presents the results of a small-scale study based on the question: What effect does a short-term course, based on Nickerson's anchoring and adjustment model, and Fischer's model of dynamic skill development, have on the diagnostic ability of prospective teachers about pupils' comprehension of technical systems? The effects monitored were the teachers' relative-implicit (RI), absolute-implicit (AI), relative-explicit (RE) and absolute-explicit (AE) diagnoses and their self-efficacy beliefs about technology education at primary schools.

Methodology

Participants

The participants were 17 male and 34 female students of a university of Applied Sciences who followed a study to become a teacher in primary education. Their age ranged from 20 to 49, with a mean age of 25.4 years ($sd = 6.1$). Eighteen participants followed the part-time program, which is meant for people who want to make a career switch to education. None of those participants had a technical background. All participants were in the final year of their study. In this phase of their study, they teach, under supervision, a few days a week in a primary school. For the participants, it was mandatory to follow several courses from a programme that included the current course. Therefore, it can be expected that the participants had some affinity with the subject. No post-test scores were available for two participants who did not finish the course.

Measurements, scoring and analyses

A pre-post design was used to evaluate the effects of the course on the diagnostic ability of the participants. The participants' technical knowledge was measured before the course as a covariate that might affect the participants' diagnostic ability and the course's impact on that ability (Pleasants & Olson, 2019; Sanjosé & Otero, 2021). An adapted version of the STEBI-b questionnaire was used to measure the self-efficacy beliefs of the participants before and after the course because we know that self-efficacy beliefs greatly influence teachers' teaching behaviour in general (Schipper et al., 2018; Hammack & Ivey, 2017). A diagnostic ability test was developed to determine the quality of the relative implicit

(RI) and explicit (RE) and the absolute implicit (AI) and explicit (AE) diagnoses of the participants. The absolute-explicit diagnosis was not included in the pre-test as it required knowledge of the Fischer scale, which was first introduced in the course.

Technical knowledge

The test to determine teachers' technical knowledge was based on several editions (2015, 2016 and 2017) of a national admission test for students who want to become primary school teachers and lack sufficient qualifications (Cito.nl). From these editions of the admission test, 35 items were selected with engineering-related content. All the items were multiple-choice questions with three or four answer options. For example, amperage is measured with an ammeter in a circuit with a resistor. Then another identical resistor is added to the circuit. What is the effect of the additional resistor on the current measured? The current (A) remains the same; (B) is doubled; (C) is halved; (D) is zero. Anderson's LR-test showed a good model fit ($LR = 17.831(31)$, $p = 0.961$) for the test. Latent scores were calculated with eRM. The distribution of the knowledge test results deviated from a normal distribution. Therefore, the effect of technical knowledge on diagnostic skills was determined using Rfit (Kolke & McKean, 2012), a nonparametric, rank-based regression method.

Self-efficacy beliefs

The participants' self-efficacy beliefs were determined at pre-test and post-test using an adapted version of the Science Teaching Efficacy Belief Instrument—pre-service (STEBI-b: (Riggs & Enochs, 1990; Bleicher, 2004). The adaptation consisted of replacing the references to science with references to engineering. For example, "When a student does better than usual in science, it is often because the teacher exerted a little extra effort" was replaced with "When a student does better than usual in engineering, it is often because the teacher exerted a little extra effort." The modified version of the STEBI-b had a Cronbach's alpha of 0.79. The mean scale score was calculated to indicate the participants' self-efficacy beliefs, as in Bleicher (2004), using a 1-to-5-point value assigned to the Likert scale and a reversed value for negatively formulated items.

Whether the course had a significant influence on the self-efficacy beliefs of the participants was analysed by comparing the pre-test and post-test scores for the adapted STEBI-b with the Wilcoxon matched-pair signed-rank test. The prepost correlation was considered in calculating effect size (Morris, 2008).

Diagnostic ability test

The diagnostic ability of the participants was tested by asking how they would interpret pupils' skills in constructing a technical system. Pupils' work products of two technical systems were used: the Buzz-Wire and the Stairs Marble Track. The Buzz-Wire (BW) is an electric circuit with a copper spiral and a ring with a handle. When the ring touches the spiral, it will activate a lamp and buzzer. The Stairs Marble Track (SMT) has a camshaft with eccentric wheels which support a set of bars of increasing height, each with a top that transports a marble in the direction of the roll-off point. Turning the camshaft allows the marble to roll onto the next bar and finally onto the slide that brings the marble back to the

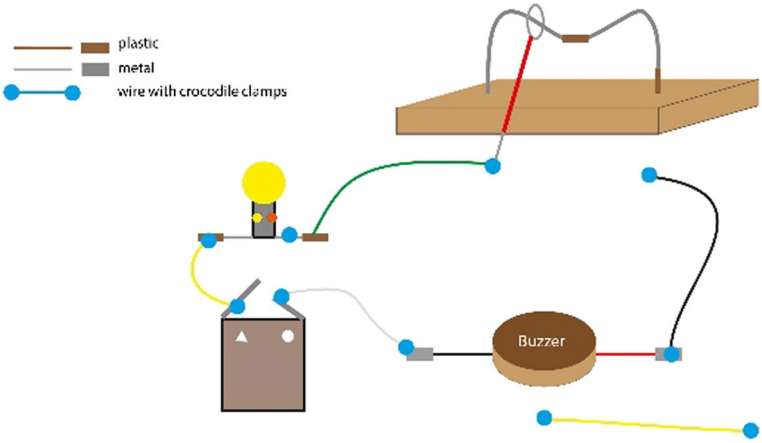

Fig. 1 Case 047 schematic drawing of a Buzz Wire work product. The pupil connects all parts. Clamps are only connected to the ends of the components' connectors (affordance), even though some of those ends are insulated (no application of knowledge about plastic being an isolator). A representation of an electrical circuit is not used

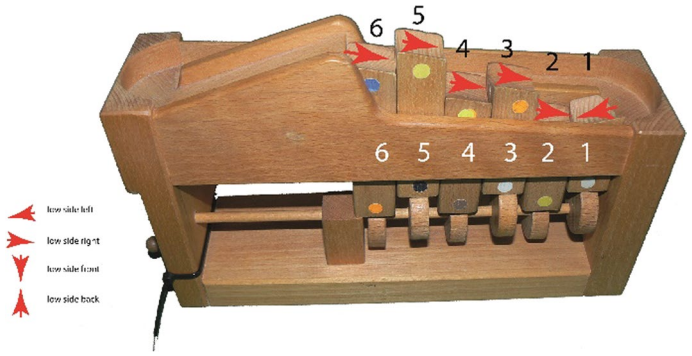

Fig. 2 Case 202 Stairs Marble Track work product, white numbers indicate correct bar positions, black numbers the bars as positioned by the pupil. The pupil applies the logic of the bar order in relation to the difference in height and the movement of the (blocked) camshaft. The slanted tops of bars 6 to 2 follow the virtual line (affordance) that connects the high roll-off point with the low roll-on point

roll-on point. Pupils from the upper primary classes were asked to construct these devices from their parts (Wammes et al., 2021). Some of these work products, representing different developmental phases of the Fischer scale, were selected for the diagnostic-ability test. A schematic drawing represented BW work products (see Fig. 1). SMT work products were represented by a photo with additional information on the bars' position (see Fig. 2). The test consisted of three pairs of BW work products and three pairs of SMT work products. For each pair, (1) a relative-implicit, (2) an absolute-implicit, (3) a relative-explicit and (4) an absolute-explicit diagnosis was required.

For the relative-implicit diagnosis (RI), the participants were asked which work product would reflect a higher level of understanding. The choice was correct when

the work product represented the higher Fischer-scale level. The percentage of correct answers per participant was calculated for both pre and post-test. The significance of the difference was calculated in R with the related sample Wilcoxon Signed Rank test.

For the absolute-implicit diagnosis (AI), the participants were asked to indicate the difference between the levels of understanding reflected by the two work products of each pair on a seven-point Likert scale. Seven points was the maximum difference between no sign of any understanding represented by a lack of reconstruction and complete understanding represented by correct reconstruction. Per participant, the Intraclass Correlation Coëfficient (ICC) (McGraw & Wong, 1996) was calculated (two-way random, single measure, consistency) for their estimates of the differences between the work products and the Fischer scale differences.

For the relative-explicit diagnosis (RE), the participants were asked to describe each pair of work products based on the knowledge they thought the pupils had applied. Two raters coded all the descriptions for three categories: system properties, technical terms and affordances. Per phrase, it was determined whether it referred to a system component or property. For example, the pupil does (not) know how to *connect* the *battery* too (..); Both pupils know how to *connect* the *lamp*; The appropriate *order of the bars* is (not) recognised; Pupil A puts the *bars upside down* on the *eccentric wheels*, fitting the *slanted top* of the bar shape onto *the rim of the eccentric wheels*. The number of unique references to a particular system feature was tallied per comparison and participant. The agreement between the raters on which phrases referred to components or system properties was high (ICC = 0.88). The number of coded system properties (RE-sys) was used to indicate the participants' ability to explain their diagnosis using the observed differences.

The use of technical terms (RE-tech) was scored because domain-specific knowledge may play a role in teachers' diagnostic ability. First, the two raters independently determined which of the terms used could be deemed to be technical. The level of agreement was high (ICC = 0.88). After consultation, a list of 'technical terms' was drawn up, which included terms like electric circuit, poles (battery), metal, insulator, conductor, camshaft and axis. The number of different technical terms used per work product description was correlated with the results of the other diagnostic measurements.

References to affordances (RE-af) were scored because affordances play a major role in pupils' understanding of systems. The role of affordances in pupils' thinking and actions was introduced in the course. The level of agreement was 0.57. The same procedure as for the technical terms was used to create a coding list of phrases that were considered as referring to affordances, like 'Pupil A makes a slide', 'Pupil A fits parts like the pieces of a puzzle', 'Pupil B uses a virtual slope (from the highest to the lowest point) to determine bar positions- and 'Both pupils consider isolation on outer points (of lamp connectors) as suitable for attaching clips'.

The significance of the difference in the sums of the participant's RE-sys, RE-tech and RE-af references between pre-test and post-test was calculated in R with the paired t-test. Cohen's *d* was calculated as an indication of the effect size.

For the absolute-explicit judgement (AE), participants were asked to determine the level of skill development for each work product using the Fischer scale-based scoring rules (see online Appendix 1). This absolute-explicit diagnosis was only solicited at the post-test because the participants did not know about the Fischer scale before entering the course. The Intraclass Correlation Coefficient (two-way mixed, absolute agreement) was used to indicate how the participants' rating matched the level calculated using an SQL version of the scoring rules (Wammes et al., 2021).

Procedure

Students took the knowledge test and the STEBI-b before the first meeting. The first meeting started with the first step of the course: becoming aware of one's perception of technical systems by constructing an electrical and a mechanical system. Then, they took the diagnostic ability pre-test that included the RI, AI and RE diagnoses. The diagnostic ability test was repeated at the end of the third meeting, with the addition of the AE diagnosis. Finally, the STEBI-b was filled-in again.

The course and the diagnostic test were piloted with six participants. The pilot resulted in some changes in the course and the diagnostic test. Some parts of the content of the course were removed, allowing more discussion about the remaining parts. Improvements were made in the wording of some of the questions and the pictures of the diagnostic test. Due to these changes, the data of these six participants were not included in the analyses. After these improvements, the course became part of the regular program. Data were collected in the first four sessions. Informed consent was requested from and given by all participants.

Results

The participants that opted for the course were interested in the domain. However, as Fig. 3 shows, with 60% correct answers on the technical knowledge test, their subject-matter knowledge for the domain was limited, which is in line with other research (Culver, 2012; Ramaligela, 2021; Sanjosé & Otero, 2021). The pre-test scores on the adapted STEBI-b ranged from 2.26 to 4.48 with an average of 3.21 ($sd = 0.43$), which is below the average scores reported for the science domain, e.g., 3.62 (Riggs & Enochs, 1990), 3.58 (Bleicher, 2004). Figure 3 shows a weak, positive, non-significant correlation between the percentage of correct answers on the knowledge test and the participants' score on the adapted STEBI-b, $r = 0.259$, $p = 0.072$).

Diagnostic ability

Relative-implicit diagnoses (RI)

For the relative-implicit diagnoses, a pair-wise judgement was used. The participants were asked to select the best from two work product pictures. No additional information was provided. In the pre-test, six participants were always correct in their choice. At post-test, thirteen participants were always correct (see Table 1). The amount of correct choices improved from 81% at pre-test (76% BW, 86% SMT) to 90% at post-test (83% BW, 96% SMT). The related-samples Wilcoxon Signed Rank Test showed a significant difference ($V = 323$, $p = 0{,}001$) with a moderate effect size ($r = 0.46$). There was no significant improvement in correct choices for BW work products ($V = 187$, $p = 0.127$), but a significant improvement for SMT work products ($V = 93$, $p = 0.010$). It can be concluded that already at the pretest, for most comparisons, the participants could identify which work product did reflect a higher level of understanding. The course did improve this ability, but this improvement was only significant for the SMT work products.

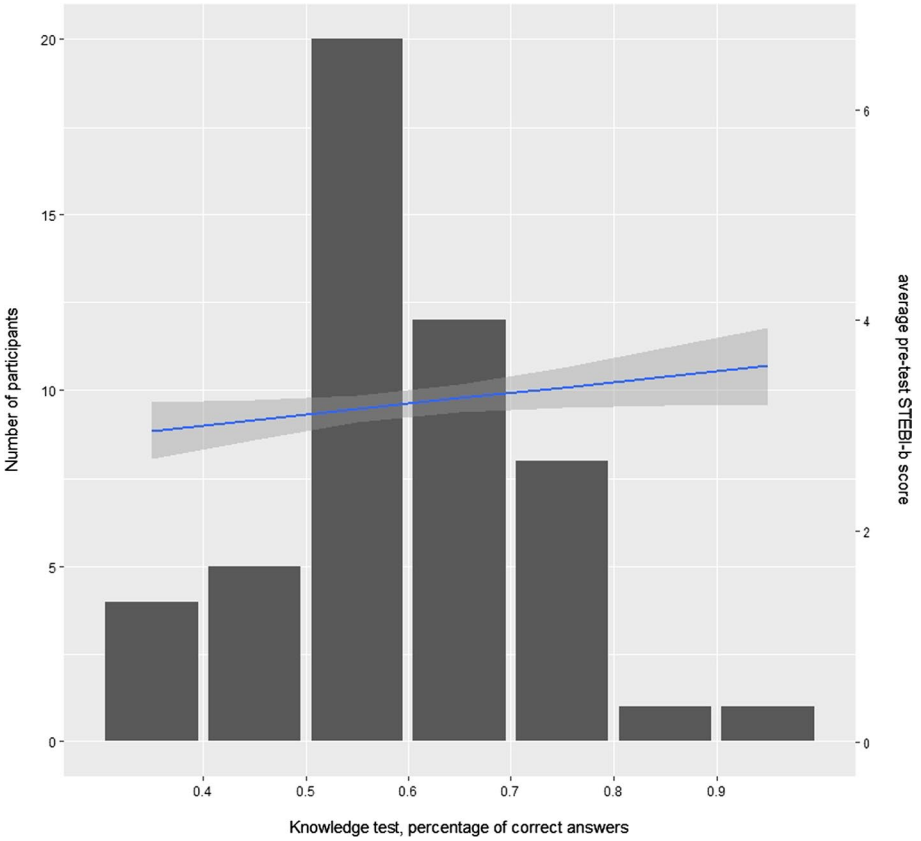

Fig. 3 Average pre-test STEBI-b score related to the knowledge test score ($n = 34$)

Table 1 Participants' relative-implicit judgements about the best out of two work products[a]

Correct	Pre-test ($n = 48$)	Post-test ($n = 46$)
All six pair-wise judgements[b]	8	17
Five pairs	24	26
Four pairs or fewer	16	3

[a]Three judgements on the best out of two BW work products and three judgements on the best out of two SMT work products

[b]Correct when the participants' choice matches the difference as determined by the Fischer-scale-based scoring rules

Absolute-implicit diagnoses

For the absolute-implicit diagnosis, the participants were asked to express the magnitude of the difference in understanding seen in the work products of each pair. At the pre-test, the correlation of their estimate with the Fischer-scale difference ranged from $ICC_{(C,1)} - 0.358$ to 0.948 with a mean of 0.547. At the post-test, these correlations ranged from − 0.517 to

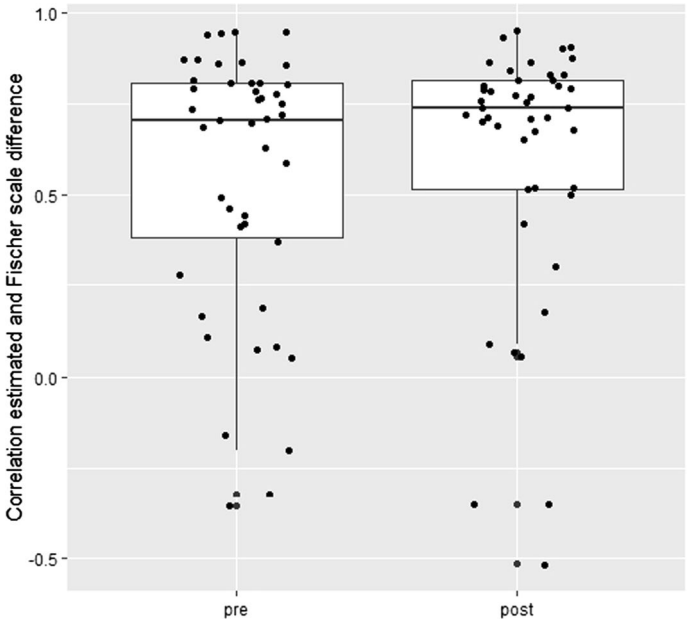

Fig. 4 Correlation between the participants' estimates of work product differences and the differences indicated by the Fischer scale ($n=46$)

0.950, with a mean of 0.603. The boxplot of Fig. 4 shows an overall improvement of the correlation between the estimates and the Fischer scale differences, but with considerable individual differences. At post-test, there were 32 participants whose estimates were in line with the Fischer scale differences, of whom nine had low correlations at pretest. Six participants regressed to a substantially lower correlation at post-test, and eight showed low correlations at both pre and post-test. There were missing values for three of the participants. The AI_{pre}–AI_{post} correlation was $r_s=0.480$, $p=0.001$.

An effect size of the course on the participants' ability to estimate differences between work products on a Likert scale could not be calculated due to the large standard errors of the correlations, which relate to the limited number of six comparisons per participant. From Fig. 4, it can be concluded that the impact of the course on the absolute-implicit estimates was positive but limited.

Relative-explicit diagnosis

In their relative-explicit diagnoses, the participants used the differences between each pair of work products to infer and describe which knowledge had been used by both pupils. The system features, technical terms, and references to affordances were coded and counted for these descriptions. At post-test, there was a significant increase in the number of described system features, $t(45)=6.413$, $p<0.001$, $d=1.04$, technical terms $t(45)=3.099$, $p=0.003$, $d=0.58$ and references to affordances, $t(45)=4.882$, $p<0.001$, $d=0.67$. Table 2 summarises the differences between pre-test and post-test for BW and SMT comparisons.

Table 2 References to system properties, affordances and technical terms in descriptions per device. Average per participant (n = 48)

	System properties (RE-sys)				Affordances (RE-af)				Technical terms (RE-tech)			
	BW		SMT		BW		SMT		BW		SMT	
	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd	Mean	sd
Pre-test	7.5	2.4	6.2	2.4	1.2	1.3	2.4	2.2	3.8	2.9	0.4	.8
Post-test	9.0	2.6	9.2	2.4	1.9	1.8	3.6	2.0	5.1	3.3	1.4	1.7

The results show that the participants recognise the role of affordances especially in the SMT, which is obvious because the SMT system is based on differences in shape. The increase in affordances mentioned in the post-test can be attributed to the course, which made the participants aware of the role of affordances in pupils' actions. The more substantial increase in system properties for the SMT may relate to the fact that they were probably more experienced with simple electric circuits like the BW task than with a mechanical system as presented by the SMT task. The relatively large standard deviations indicate substantial differences between the participants, not only in their ability to recognise the differences but also in how they answered the question. Some participants answered in keywords and emphasised the most obvious differences, while others provided their answers in full sentences and mentioned both differences and similarities. The effect sizes indicate that the course improved the ability of the participants to notice and express the differences in terms of system features, technical terms, and references to affordances between the work products.

Absolute-explicit diagnosis

The absolute-explicit diagnosis required the application of the Fischer scale, which was introduced in the course. Therefore, it was only asked in the post-test, where the participants rated the six SMT and six BW work products using Fischer scale-based scoring rules. Figure 5 provides an overview of the distribution of the participants' ICC scores, which indicate the agreement rate between the participants' estimates of pupils' level on the Fischer scale and the scale level based on the heuristics published by Wammes et al. (2021). Out of 47 participants, 32 had an ICC above 0.7.

The participants' ratings rarely deviated by more than one level from the algorithm-based reference (see Fig. 6). It may be concluded that after the course, most participants were able to use the Fischer scale to interpret the developmental level reflected by pupils' work products.

Participants' diagnostic ability

The results presented above indicate a positive effect of the course. However, not every participant benefited, and not all participants showed progress on all types of diagnoses. Table 3 shows the participants' post-test performances compared to the pre-test.

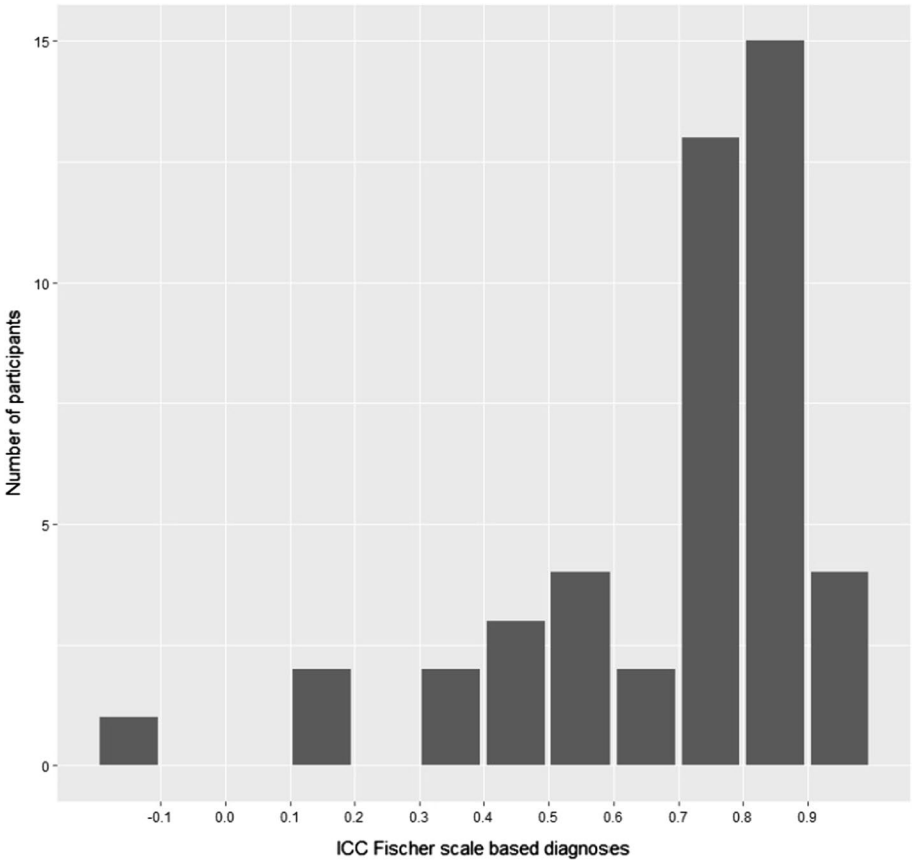

Fig. 5 Distribution of ICC scores based on six BW and six SMT Fischer scale ratings ($n = 47$)

There were significant post-test correlations between the correct application of the scoring rules (AE) and the relative (RI) and absolute implicit diagnoses (AI) of the participants (see Table 4). Except for (RE_tech) and (AI), there was a lack of correlation between the diagnostic parameters (AE, RI and AI) and which differences in pupils' knowledge the participants identified.

Rfit showed that the results of the knowledge test had a positive but non significant effect on RE-sys_{post} ($b = 7.78$, $t = 1.20$, $p = 0.24$), RE-$tech_{post}$ ($b = 9.09$, $t = 1.57$, $p = 0.12$), RE-af ($b = 7.50$, $t = 1.64$, $p = 0.11$) and AE, the correct application of scoring rules ($b = 0.13$, $t = 0.72$, $p = 0.47$). No effect was found on AI_{post} ($b = -0.12$, $t = -0.476$, $p = 0.64$). This implies that the ability to identify differences has only a weak relationship with technical knowledge.

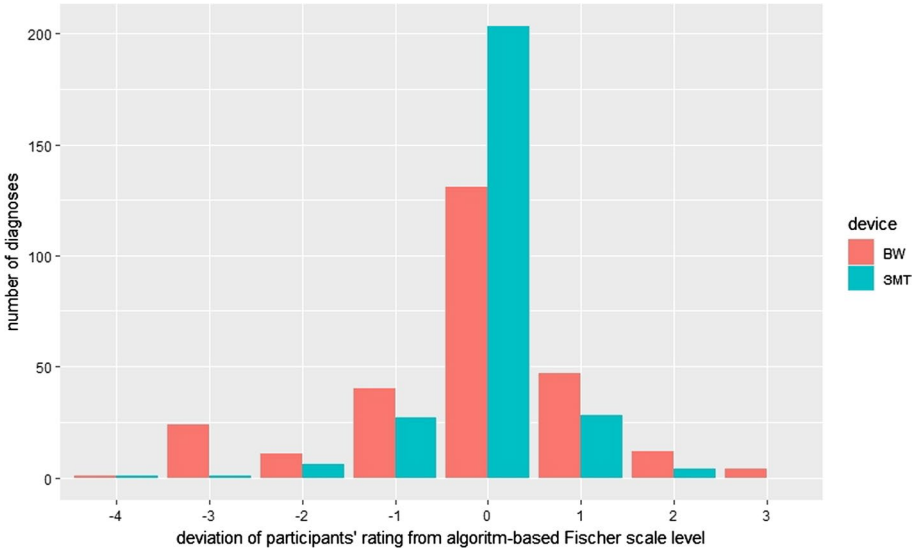

Fig. 6 Participants' Fischer scale ratings compared to an algorithm-based rating ($n = 47$)

Table 3 Performance of participants by test compared to their pre-test result (n = 47)

Type of diagnosis	Decline	Equivalent	Improvement
Relative-implicit accuracy (RI)	6	18	22
Absolute-implicit accuracy (AI)	6	23	15
References to system properties (RE_sys)	4	13	29
References to affordances (RE_af)	4	16	26
Use of technical vocabulary (RE_tech)	7	14	25

Table 4 Spearman correlation post-test measurements

	RE_tech	RE_sys	RE_af	AE	AI	RI
RE_tech		.419**	− .171	.216	.313*	− .135
RE_sys			− .061	.009	− .079	− .135
RE_af				− .256	− .083	.007
AE					.504*	.291*
AI						.101
RI						

**Significant at the .01 level, *Significant at the .05 level (2-tailed)

Self-efficacy beliefs

At post-test, participants' self-efficacy beliefs significantly increased from 3.24 to 3.60 on the five-point scale, $t(42) = 7.83$, $p < 0.001$. The effect size d was 1.19, 95%

CI [0.80; 1.58]. The STEBI-b showed a weak positive non-significant correlation with the knowledge test results at pre-test ($r = 0.253$, $p = 0.072$) and at post-test ($r = 0.256$, $p = 0.086$).

Discussion

Diagnostic skills enable teachers to adjust their instruction, tasks and feedback to their pupils' prior knowledge. It has been demonstrated that most teachers in primary education have limited diagnostic skills for technology education. To improve these skills, we developed a course for prospective teachers that consisted of three 90 min sessions, including about 50 min of testing. Like Ostermann et al., we used Nickerson's (1999) anchoring and adjustment model for the course design. This model aims to create awareness about the differences between the participants' ways of thinking about technical systems and pupils' ways of thinking and reactions in subsequent developmental phases. We estimated the effects of the course on the teachers' diagnostic ability by their relative-implicit (RI), absolute-implicit (AI), relative-explicit (RE) and absolute-explicit (AE) judgements and their self-efficacy beliefs about technology education at primary schools.

Most relative implicit (RI) judgements were already correct at the pre-test. At post-test, all judgements about SMT pairs were correct, and overall there was a significant improvement with a moderate effect size.

The absolute implicit (AI) judgements were less accurate and showed large differences between teachers. At the pretest, about half of the teachers estimated the difference in knowledge application in line with the difference indicated by the Fischer scale. At post-test, about two-thirds of the teachers could infer the differences in understanding reflected by pupils' work products. Still, one-third of the teachers lacked accuracy for such estimates. The significant correlation with the AE post-test results suggests that those who could apply the Fischer scale could also make good estimates of the differences before determining the work products' position on the Fischer scale. However, the strong correlation between AI_{pre} and AI_{post} opposes the suggestion that introducing the Fischer scale in the course explains the high AI_{post} scores of two-thirds of the participants.

The course had a significant positive effect on the teachers' ability to express the differences and similarities between the work products (RE) in terms of system properties (RE-sys), the use of technical terms (RE-tech) and their description of the possible influence of affordances (RE-af). At post-test, the teachers' diagnoses of the Fischer-scale level of the work products (AE) were in line with the levels calculated by the research team. There was a remarkably strong effect on the teachers' self-efficacy scores. Overall it may be concluded that the course supported most participants in developing their ability to diagnose pupils' understanding of technical systems. This supports the finding of Ostermann et al. (2018) that Nickerson's model (1999) might be an appropriate structure for a course to improve the diagnostic ability of pre-service teachers. More emphasis on the course as a way to become aware of how the unconscious application of one's thinking in instruction and feedback might sometimes be ineffective in bringing pupils' thinking forward might even attract more female prospective teachers to follow a course with scientific and technical content.

The results align with previous findings that teachers are generally better in their relative than their absolute estimates of performance (Lesterhuis et al., 2017; Südkamp et al., 2012). According to Schrader and Helmke (2001), an explanation of this difference is that a ranking better reflects the individual teacher's perspective on student performance. The significant correlations between the correct application of the scoring rules (AE) and the RI and AI judgements at post-test might indicate that those who understood the Fischer scale did apply that knowledge in their intuitive estimates. This might indicate that learning about and practising with the Fischer scale has improved the participants' ability to notice differences in pupils' technical ability.

The limited and sometimes negative correlations between RI, AI and AE, on the one hand, and the RE scores, on the other hand, imply that teachers who can indicate which work product is the better one who can provide a good estimate of the magnitude of the differences and who can correctly apply scoring rules are not necessarily the teachers who can express the differences. That implies that a course to improve a teacher's diagnostic ability should not only focus on diagnostic accuracy but also on the ability to explain why something does not meet all the demands of a properly working system. Such an ability is crucial to provide tailored feedback that can help pupils' understanding forward (Van de Pol et al., 2010).

The average technological knowledge of the participants was about 60%, which differs from the average of 80% found by Rohaan (2012). Rohaan et al. noticed that their test was probably too easy for pre-service teachers, as their questions were originally constructed for sixth-graders (age 11–12). The current test seems better suited to revealing differences in technological knowledge between pre-service teachers. Sufficient subject matter knowledge has been identified as important for the quality of instruction (Hartell et al. 2015; Jones & Compton, 1998; Pleasants & Olson, 2019; Rohaan et al., 2009; Utley et al., 2019). For diagnostic ability, this study showed positive but non-significant correlations between the participant's scores on the knowledge test and their diagnostic ability. Therefore, it can be concluded that the participants' technical knowledge might influence their diagnostic capabilities but did not determine their scores on the diagnostic ability tests used in this study.

The strong positive effect of the course on the self-efficacy beliefs of the participants was an important side effect, as greater self-confidence is positively related to dedicating time to engineering education (Van Cleynenbreugel et al., 2011; Van der Molen, 2008). There are indications that experience improves the diagnostic ability of teachers (Wammes et al., 2022). Strong self-efficacy beliefs might support novice teachers to get experienced in teaching science and technology.

Limitations

An important limitation of the present study is the limited number of participants combined with the elective nature of the course. Therefore, it cannot be assumed that the course would generate similar results with a random group of prospective or in-service teachers. The focus on two technical systems is another limitation. Engineering is a multi-faceted domain with specific skills (Mitcham, 1994; Pearson & Young, 2002). It is not likely that the trained teachers would be able to apply the knowledge about skill development to other

types of systems or technical skills (Perkins & Salomon, 1992). Therefore, the conclusions about the effects are limited to the two technical systems used in the course.

According to Nickerson's (1999) model, insight into the thinking of others primarily arises from an awareness of how one's knowledge differs from that of others. This study did not explicitly examine the participants' or pupils' thinking about technical systems. Thus, the conclusion about the effectivity of this model should be considered as hypotheses based on observations of the results of pupils' and participants' considerations. Additionally, it should be emphasised that much technical knowledge is tacit and therefore offers limited opportunities for direct assessment.

Finally, strengthening diagnostic capacity does not necessarily equate to enabling teachers to adapt their instruction and feedback optimally to observed skill-level differences. To what extent that requires additional training in other aspects of teachers' pedagogical content knowledge require further research.

Implications and conclusion

Learning outcomes improve when teachers can adapt their instruction and feedback to differences in the proficiency levels of their pupils. This requires the ability to identify such levels through pupils' behaviour. It became clear that most teachers differed in their ability to compare pupils' results intuitively, analyse pupils' thinking, and interpret pupils' results using abstract scoring rules. This implies that the effect of a course on teachers' diagnostic ability should not be pinned down to a single variable. It can be concluded that a course based on Nickerson's model can positively affect the participants' diagnostic abilities and their self-efficacy beliefs about technology education.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10798-022-09774-x.

References

Assaraf, O. B., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching, 47*(5), 540–563. https://doi.org/10.1002/tea.20351

Bassano, D., & VanGeert, P. (2007). Modelling continuity and discontinuity in utterance length: A quantitative approach to changes, transitions and intra-individual variability in early grammatical development. *Developmental Science, 10*(5), 588–612. https://doi.org/10.1111/j.1467-7687.2007.00629.x

Behrmann, L., & Souvignier, E. (2013). Pedagogical content beliefs about reading instruction and their relation to gains in student achievement. *European Journal of Psychology of Education, 28*, 1023–1044. https://doi.org/10.1007/s10212-012-0152-3

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment Granada Learning.

Bleicher, R. E. (2004). Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *School Science and Mathematics, 104*(8), 383–391. https://doi.org/10.1111/j.1949-8594.2004.tb18004.x

Brophy, S., Klein, S., Portsmore, M., & Rogers, C. (2008). Advancing engineering education in P-12 classrooms. *Journal of Engineering Education, 97*(3), 369–387. https://doi.org/10.1002/j.2168-9830.2008.tb00985.x

Butler, R. (2012). Striving to connect: Extending an achievement goal approach to teacher motivation to include relational goals for teaching. *Journal of Educational Psychology, 104*(3), 726–742. https://doi.org/10.1037/a0028613

Chemero, A. (2003). An outline of a theory of affordances. *Ecological Psychology, 15*(2), 181–195. https://doi.org/10.1207/S15326969ECO1502_5

Culver, D. E. (2012). A qualitative assessment of preservice elementary teachers' formative perceptions regarding engineering and K-12 engineering education.

DePaulo, B. M., Charlton, K., Cooper, H., Lindsay, J. J., & Muhlenbruck, L. (1997). The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*(4), 346–357. https://doi.org/10.1207/s15327957pspr0104_5

Dunlosky, J., & Thiede, K. W. (2013). Four cornerstones of calibration research: Why understanding students' judgements can improve their achievement. *Learning and Instruction, 24*, 58–61. https://doi.org/10.1016/j.learninstruc.2012.05.002

Edelenbos, P., & Kubanek-German, A. (2004). Teacher assessment: The concept of 'diagnostic competence.' *Language Testing, 21*(3), 259–283.

Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of the mind*. Basic Books.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*(6), 477. https://doi.org/10.1037/0033-295X.87.6.477

Fischer, K. W., & Bidell, T. R. (2007). Dynamic development of action and thought. *Handbook of Child Psychology*. https://doi.org/10.1002/9780470147658.chpsy0107

Gibson, J. J. (1977). *The theory of affordances*. Hilldale.

Gingerich, A., Kogan, J., Yeates, P., Govaerts, M., & Holmboe, E. (2014). Seeing the 'black box' differently: Assessor cognition from three research perspectives. *Medical Education, 48*(11), 1055–1068. https://doi.org/10.1111/medu.12546

Govaerts, M., Van de Wiel, M., Schuwirth, L., Van der Vleuten, C., & Muijtjens, A. (2013). Workplace-based assessment: Raters' performance theories and constructs. *Advances in Health Sciences Education, 18*(3), 375–396. https://doi.org/10.1007/s10459-012-9376-x

Hammack, R., & Ivey, T. (2017). Examining elementary teachers' engineering self-efficacy and engineering teacher efficacy. *School Science and Mathematics, 117*(1–2), 52–62. https://doi.org/10.1111/ssm.12205

Hartell, E., Gumaelius, L., & Svärdh, J. (2015). Investigating technology teachers' self-efficacy on assessment. *International Journal of Technology and Design Education, 25*(3), 321–337. https://doi.org/10.1007/s10798-014-9285-9.

Hattie, J. (2013). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hsu, M., Purzer, S., & Cardella, M. E. (2011). Elementary teachers views about teaching design, engineering, and technology. *Journal of Pre-College Engineering Education Research (j-PEER), 1*(2), 5. https://doi.org/10.5703/1288284314639

Jones, A., & Compton, V. (1998). Towards a model for teacher development in technology education: From research to practice. *International Journal of Technology and Design Education, 8*(1), 51–65. https://doi.org/10.1023/A:1008891628375

Jones, A., & Moreland, J. (2004). Enhancing practising primary school teachers' pedagogical content knowledge in technology. *International Journal of Technology and Design Education, 14*(2), 121–140. https://doi.org/10.1023/B:ITDE.0000026513.48316.39

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kaiser, J., Helm, F., Retelsdorf, J., Südkamp, A., & Möller, J. (2012). Zum Zusammenhang von Intelligenz und Urteilsgenauigkeit bei der Beurteilung von Schülerleistungen im Simulierten Klassenraum [On the relation of intelligence and judgement accuracy in the process of assessing student achievement in the simulated classroom]. *Zeitschrift Für Pädagogische Psychologie / German Journal of Educational Psychology, 26*(4), 251–261. https://doi.org/10.1024/1010-0652/a000076

Kloke, J. D., & McKean, J. W. (2012). Rfit: Rank-based estimation for linear models. *R J., 4*(2), 57.

Knezek, G., Christensen, R., & Tyler-Wood, T. (2011). Contrasts in teacher and student perceptions of STEM content and careers. *Contemporary Issues in Technology and Teacher Education, 11*(1), 92–117.

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Maeyer, S. (2017). Comparative judgement as a promising alternative to score competencies. In *Innovative practices for higher education assessment and measurement* (pp. 119–138). IGI Global. https://doi.org/10.4018/978-1-5225-0531-0.ch007

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30. https://doi.org/10.1037/1082-989X.1.1.30

Meindertsma, H. B., van Dijk, M. W. G., Steenbeek, H. W., et al. (2014). Assessment of preschooler's scientific reasoning in adult-child interactions: What is the optimal context? *Research in Science Education, 44*(2), 215–237. https://doi.org/10.1007/s11165-013-9380-z

Mitcham, C. (1994). *Thinking through technology: The path between engineering and philosophy*. University of Chicago Press.

Molnár, G., Greiff, S., Wustenberg, S., & Fischer, A. (2017). *Empirical study of computer based assessment of complex problem solving skills*. Paris: OECD Publising. https://doi.org/10.1787/9789264273955-en

Moreland, J., & Jones, A. (2000). Emerging assessment practices in an emergent curriculum: Implications for technology. *International Journal of Technology and Design Education, 10*(3), 283–305. https://doi.org/10.1023/A:1008990307060

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*(2), 364–386. https://doi.org/10.1177/1094428106291059

Mullis, I. V., & Martin, M. O. (2017). *TIMSS 2019 assessment frameworks*. TIMMS & PIRLS International Study Center, Lynch School of Education, Boston College.

National Assessment Governing Board. (2013). *Technology and engineering literacy framework*. Reingold Inc. http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/tel-abridged-2014.pdf

Nadelson, L. S., Callahan, J., Pyke, P., Hay, A., Dance, M., & Pfiester, J. (2013). Teacher STEM perception and preparation: Inquiry-based STEM professional development for elementary teachers. *The Journal of Educational Research, 106*(2), 157–168. https://doi.org/10.1080/00220671.2012.667014

Nauta, M. M., Kahn, J. H., Angell, J. W., & Cantarelli, E. A. (2002). Identifying the antecedent in the relation between career interests and self-efficacy: Is it one, the other, or both? *Journal of Counseling Psychology, 49*(3), 290–301. https://doi.org/10.1037/0022-0167.49.3.290

Nickerson, R. S. (1999). How we know—And sometimes misjudge—What others know: Imputing one's own knowledge to others. *Psychological Bulletin, 125*(6), 737–759. https://doi.org/10.1037/0033-2909.125.6.737

Ostermann, A., Leuders, T., & Nückles, M. (2018). Improving the judgement of task difficulties: Prospective teachers' diagnostic competence in the area of functions and graphs. *Journal of Mathematics Teacher Education, 21*, 579–605. https://doi.org/10.1007/s10857-017-9369-z

Pearson, G., & Young, A. T. (2002). *Technically speaking: Why all Americans need to know more about technology*. National Academies Press.

Perkins, D. N., & Salomon, G. (1992). Transfer of learning. *International Encyclopedia of Education, 2*, 6452–6457.

Pleasants, J., & Olson, J. K. (2019). Refining an instrument and studying elementary teachers' understanding of the scope of engineering. *Journal of Pre-College Engineering Education Research (j-PEER), 9*(2), 1. https://doi.org/10.7771/2157-9288.1207

Ramaligela, S. M. (2021). Exploring pre-service technology teachers' content and instructional knowledge to determine teaching readiness. *International Journal of Technology and Design Education, 31*(3), 531–544.

Ready, D. D., & Wright, D. L. (2011). Accuracy and inaccuracy in teachers' perceptions of young children's cognitive abilities: The role of child background and classroom context. *American Educational Research Journal, 48*(2), 335–360. https://doi.org/10.3102/0002831210374874

Retnawati, H., Djidu, H., Kartianom, A., & Anazifa, R. D. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in the 21st Century, 76*(2), 215–230.

Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education, 74*(6), 625–637.

Rohaan, E. J., Taconis, R., & Jochems, W. M. (2009). Measuring teachers' pedagogical content knowledge in primary technology education. *Research in Science & Technological Education, 27*(3), 327–338. https://doi.org/10.1080/02635140903162652.

Rohaan, E. J., Taconis, R., & Jochems, W. M. (2012). Analysing teacher knowledge for technology education in primary schools. *International Journal of Technology and Design Education, 22*(3), 271–280. https://doi.org/10.1080/02635140903162652

Sanjosé, V., & Otero, J. (2021). Elementary pre-service teachers' conscious lack of knowledge about technical artefacts. *International Journal of Technology and Design Education*. https://doi.org/10.1007/s10798-021-09696-0

Scharten, R., & Kat-deJong, M. (2012). *Koersvast en enthousiast kritieke succesfactoren van gelderse vindplaatsen [Enthusiastic and purposeful. What makes primary schools in Gelderland succesful in their science and technology education]*. Expertisecentrum Nederlands.

Schrader, F., & Helmke, A. (2001). Alltägliche leistungsbeurteilung durch lehrer [Everyday performance appraisal by teachers]. In F. E. Weinert (Ed.), *Leistungsmessungen in schulen [Performance measurements in schools]* (pp. 45–58).

Schipper, T., Goei, S. L., de Vries, S., & van Veen, K. (2018). Developing teachers' self-efficacy and adaptive teaching behaviour through lesson study. *International Journal of Educational Research, 88*, 109–120. https://doi.org/10.1016/j.ijer.2018.01.011.

Shavelson, R. J. (1978). Teachers' estimates of students' states of mind and behavior. *Journal of Teacher Education, 29*(5), 37–40. https://doi.org/10.1177/002248717802900511

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgements of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743–762. https://doi.org/10.1037/a0027627

Svensson, M., Zetterqvist, A., & Ingerman, Å. (2012). On young people's experience of systems in technology. *Design & Technology Education: An Internatioal Journal, 17*(1).

Sweeney, L. B., & Sterman, J. D. (2007). Thinking about systems: Student and teacher conceptions of natural and social systems. *System Dynamics Review, 23*(2–3), 285–311. https://doi.org/10.1002/sdr.366

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. MIT Press.

Urhahne, D., & Wijnia, L. (2021). A review on the accuracy of teacher judgements. *Educational Research Review, 32*, 100374. https://doi.org/10.1016/j.edurev.2020.100374

Utley, J., Ivey, T., Hammack, R., & High, K. (2019). Enhancing engineering education in the elementary school. *School Science and Mathematics, 119*(4), 203–212. https://doi.org/10.1111/ssm.12332

Van Cleynenbreugel, C., De Winter, V., Buyse, E., & Laevers, F. (2011). Understanding the physical world: Teacher and pupil attitudes towards science and technology. In *Professional development for primary teachers in science and technology* (pp. 121–143). Springer.

Van de Pol, J., Volman, M., & Beishuizen, J. (2010). Scaffolding in teacher–student interaction: A decade of research. *Educational Psychology Review, 22*(3), 271–296. https://doi.org/10.1007/s10648-010-9127-6

Van der Steen, S. (2014). "How does it work?": A longitudinal microgenetic study on the development of young children's understanding of scientific concepts. (Doctoral dissertation). Retrieved from http://hdl.handle.net/11370/408b8e4e-2be4-4312-a48a-8898995dc273

Walma van der Molen, J. (2008). De belangstelling voor wetenschap en techniek in het basisonderwijs [The interest in science and technology in primary education]. In D. Fourage, & A. de Grip (Eds.), *Technotopics III: Essays over onderwijs en arbeidsmarkt voor bètatechnici [Technotopics III: Essays on education and the employment market for hard science technicians]* (pp. 12–21). Den Haag: Platform Bèta Techniek. Retrieved from http://dare.uva.nl/record/306426

Wammes, D., Slof, B., Schot, W., & Kester, L. (2021). Pupils' prior knowledge about technological systems: Design and validation of a diagnostic tool for primary school teachers. *International Journal of Technology and Design Education*, 1–33.

Wammes, D., Slof, B., Schot, W., & Kester, L. (2022). Teacher judgement accuracy of technical abilities in primary education. *International Journal of Technology and Design Education*, 1–35.

Wood, T. J. (2014). Exploring the role of first impressions in rater-based assessments. *Advances in Health Sciences Education, 19*(3), 409–427. https://doi.org/10.1007/s10459-013-9453-9