

Metacognitive confidence judgments and their link to complex problem solving



Julia Rudolph^{a,*}, Christoph Niepel^a, Samuel Greiff^{a,1}, Frank Goldhammer^b, Stephan Kröner^c

^a University of Luxembourg, Luxembourg

^b German Institute for International Educational Research and Centre for International Student Assessment (ZIB), Germany

^c Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

ARTICLE INFO

Keywords:

Response confidence
Complex problem solving
Reasoning
Metacognition

ABSTRACT

With the aim to better understand the nature of complex problem solving (CPS), we investigated the link between confidence judgments, which represent a major constituent of metacognitive self-monitoring, and CPS by regressing the two facets of CPS (i.e., knowledge acquisition and knowledge application) on confidence in CPS. To ensure that the link between confidence in CPS and CPS is distinct, we controlled for reasoning, which is the strongest known correlate of CPS. Using structural equation modeling in a sample of 471 German seventh-grade students, we found that confidence in CPS explained 67% of the variance in CPS knowledge acquisition and 55% of the variance in CPS knowledge application. These links were reduced but remained substantial when we controlled for reasoning. The results indicate that confidence judgments as indicators of metacognitive monitoring in CPS are substantially linked to successful CPS, thus bringing us one step closer to a full understanding of CPS.

1. Introduction

Complex problem solving (CPS) describes the process of successfully interacting with nonroutinely encountered and dynamically changing environments (Buchner, 1995). As such, CPS is needed in a variety of situations that virtually all students face in their everyday lives. For instance, CPS is required when students use new technological devices (smartphones, computer programs, etc.; Wüstenberg, Greiff, & Funke, 2012) and is needed to cope with everyday life in professional, private, and educational settings. Its relevance has been recognized in educational research (Greiff et al., 2013), and a large body of research has recently emerged to try to understand the nature of CPS by focusing on the relation between CPS and concurrent cognitive skills (Kröner, Plass, & Leutner, 2005). One overarching finding is that CPS overlaps substantially with reasoning. However, reasoning does not account for all of the variance in CPS (e.g., see the meta-analysis by Stadler, Becker, Gödker, Leutner, & Greiff, 2015). This coincides with CPS theory, which proposes that constructs other than reasoning play important roles in CPS. In particular, self-regulation, the skill that is used to monitor and adapt one's problem-solving strategies, should be a central component of CPS (Funke, 2003; Ifenthaler, 2012; Mayer, 1998; Wüstenberg, Stadler, Hautamäki, & Greiff, 2014). In CPS, it is likely that students

do not apply expedient systematic approaches for investigating dynamic problem characteristics from the very beginning. Thus, they need to realize that they have to continuously adapt their problem-solving behavior to be successful in CPS.

Among the core aspects of self-regulation is metacognitive self-monitoring, that is, the observing and judging of one's own performance (Fritzsche, Kröner, Dresel, Kopp, & Martschinke, 2012; Stankov, 2000; Stankov, Lee, Luo, & Hogan, 2012). Both performance judgments and self-reported confidence have been shown to be both economic and valid indicators of metacognitive self-monitoring (Fritzsche, et al., 2012; Händel, & Fritzsche, 2016; Stankov, 2000).

In the present study, we investigated whether and to what extent metacognitive monitoring in complex problem solving is linked to overall performance in CPS. Moreover, we aimed to take the first steps toward ensuring that such a link cannot be ascribed to reasoning, which is at the core of intelligence (Carroll, 2003) and is a strong and well-known correlate of CPS (see the meta-analysis by Stadler et al., 2015).

1.1. Confidence in CPS and CPS

In more detail, CPS reflects the skill needed to understand and to subsequently control dynamic environments that require the problem

* Corresponding author at: University of Luxembourg, Maison des Sciences Humaines, 11, Porte des Sciences, 4366 Esch-Belval, Luxembourg.

E-mail address: julia.rudolph@uni.lu (J. Rudolph).

¹ Samuel Greiff is one of two authors of the commercially available COMPRO-test that is based on the multiple complex systems approach and that employs the same albeit very general assessment principle as MicroDYN. However, for any research and educational purpose, a free version of MicroDYN is available. Samuel Greiff receives royalty fees for COMPRO.

solver to actively interact with this environment (Buchner, 1995) in a situation in which not all the information that is needed to control the environment is apparent from the beginning and has to be discovered through active interaction (Funke, 2001). CPS research focuses on two central and intertwined facets of CPS: knowledge acquisition and knowledge application (e.g., Greiff, Kretzschmar, Müller, Spinath, & Martin, 2014; Greiff et al., 2013; Kröner et al., 2005). Successful knowledge acquisition is a skill that enables people to explore and understand the functioning of a complex system, whereas knowledge application is needed to be able to use one's understanding of the environment to control it. To give an example, if students have to edit pictures with an unfamiliar picture editor, they need to explore and understand the functions of the editor (knowledge acquisition) and subsequently use the acquired knowledge to control the editor (knowledge application).

Research that is aimed at understanding CPS has had a strong focus on the relations between CPS and other cognitive ability measures with a particular focus on reasoning because reasoning is a central component of intelligence (Carroll, 2003). One often-reported finding is that CPS overlaps substantially with cognitive ability measures, in particular with reasoning (Kröner, et al., 2005), whereas there is also some evidence for the unique explanatory value of CPS regarding external outcomes such as academic achievement (e.g., Greiff et al., 2013; Wüstenberg et al., 2012; cf. Kretzschmar, Neubert, Wüstenberg, & Greiff, 2016). However, numerous scholars have noted that other constructs may also be important for CPS, including noncognitive aspects. For instance, CPS is supposed to be highly relevant for metacognition (e.g., Funke, 2003; Ifenthaler, 2012; Huber, 1995; Mayer, 1998), which is the “cognition that reflects on, monitors, or regulates first order cognition” (Kuhn, 2000, p. 178; see also Flavell, 1979; Kleitman, & Gibson, 2011; Schraw, Crippen, & Hartley, 2006). Conceptually, metacognition is intertwined with skill in revising and evaluating tasks, goals, and strategies (Flavell, 1979). As problem solving requires such metacognitive skills to apply or adapt problem-solving strategies whenever necessary (Ifenthaler, 2012), it is conceptually linked to metacognition. Moreover, complex problems may particularly require metacognitive self-monitoring skills (Vollmeyer & Rheinberg, 1999) because students might be overwhelmed by the quantity of information that has to be tackled (Zimmermann & Campillo, 2003).

As metacognition requires skill in evaluating and adapting the process of problem solving, one aspect that is at the core of metacognition is one's judgment of one's own performance (Stankov, 2000). Throughout a large part of the literature, confidence is defined as “a state of being certain about the success of a particular behavioral act” (Stankov et al., 2012, p. 1). That is, it may be considered a combination of (a) a performance judgment and (b) certainty about this judgment. To disentangle these two aspects, in the present paper, we focused on the first aspect of confidence. For instance, during each CPS task we assessed the students' confidence by asking them whether they thought they were solving it correctly. This aspect of confidence should be a central aspect of CPS because CPS constantly requires accurate performance judgments (Funke, 2003; Greiff & Fischer, 2013; Huber, 1995), and effective problem solvers should evaluate their performance during problem solving to decide whether they should continue following the strategy they chose or whether an adaptation is needed (Funke, 2003; Mayer, 1998). Thus, the correct estimation of their performance should foster successful CPS.

At the same time, confidence may be influenced by students' CPS performance. For example, if students are aware that they are usually successful in problem solving, they may be more likely to provide positive performance judgments during problem solving tasks. This is particularly pertinent to CPS (as opposed to static problem solving) because CPS tasks offer a large number of task-inherent cues due to their dynamic and interactive character: Students may be aware that they are on the right track on the basis of task-inherent cues

(Kröner & Biermann, 2007). For instance, if the complex problem was to attach a blue frame to a picture with an unfamiliar picture editor and students have figured out how to attach a red frame, they might be aware that they are on the right track. Moreover, complex tasks enable students to verify their acquired knowledge. For instance, if students wish to verify that they know how to attach a blue frame, they can attach the frame to another picture.

Taken together, theory suggests a strong mutual link between CPS and confidence in CPS, and this should translate into a unique statistical effect when trying to explain variance in CPS performance with confidence in CPS. Whereas such a link between confidence and CPS has not been investigated empirically, confidence in cognitive tests has been shown to be substantially associated with other skills such as reasoning and perception (Kröner & Biermann, 2007) as well as writing, speaking, and numeracy (Stankov & Lee, 2008). Further, Kleitman and Gibson (2011) and Kleitman and Moscrop (2010) reported that test-specific confidence is linked to students' grades in mathematics, spelling, and reading, making it seem likely that confidence in CPS is linked to CPS performance. Altogether, we hypothesized that confidence in CPS would explain variance in the CPS facets knowledge acquisition and knowledge application as well as in CPS variance that is shared by the two facets (Hypothesis 1).

1.2. Reasoning, confidence in CPS, and CPS

In a second step, we aimed to put any gain in knowledge about the link between confidence and CPS performance (Hypothesis 1) together with the most established correlate of CPS: reasoning. As indicated above, researchers who have previously attempted to understand the nature of CPS have come to the conclusion that reasoning is a crucial constituent of CPS (see the meta-analysis by Stadler et al., 2015) because both reasoning and CPS require the abilities to identify relevant information (Babcock, 2002; Funke, 2001) and to elaborate and apply strategies to solve problems (Wüstenberg et al., 2012).

Nevertheless, reasoning and CPS differ conceptually with regard to additional requirements because all information that is needed to solve the problem is given at the outset for reasoning tasks but not for CPS tasks (Greiff et al., 2014; Wüstenberg et al., 2012; cf. Kretzschmar et al., 2016; Wittmann & Süß, 1999). Complex problems actively generate the particular situations that provide the relevant information needed to solve the problems. Thus, CPS demands an active and systematic search for missing information and the skill needed to control dynamic environments that change with time or as the problem solver interacts with them (Wüstenberg et al., 2012).

Empirically, the investigation into whether reasoning and CPS differ is often conducted by regressing the CPS facets of knowledge acquisition and knowledge application on reasoning (e.g., Greiff et al., 2014; Greiff et al., 2013; Kröner et al., 2005; Wüstenberg et al., 2012). The overall findings of such studies were that reasoning could not account for all of the variance in knowledge acquisition and knowledge application and that the two facets remained correlated when reasoning was controlled for. This indicates that the facets share CPS-specific variance that goes beyond reasoning. As Kröner (2014) stated, this variance might be partially explained by self-evaluation as a core aspect of self-regulation, which can be operationalized via CPS performance judgments. Thus, we hypothesized that confidence in CPS and CPS would be linked beyond reasoning. In particular, we expected that including confidence in CPS would explain variance in the CPS facets knowledge acquisition and knowledge application as well as CPS variance that is shared by the two facets beyond reasoning (Hypothesis 2).

1.3. Hypotheses

With the present research, we aimed to explore the nature of CPS by investigating its link to metacognitive confidence judgments in CPS. To

take the first steps toward ensuring that confidence in CPS would be found to be a distinct correlate of CPS beyond related constructs, we aimed to confirm that confidence in CPS is linked to CPS beyond reasoning, which has been shown to be a strong—if not the strongest—correlate of CPS (Greiff et al., 2014; Greiff et al., 2013; Kröner et al., 2005; Wüstenberg et al., 2012).²

We aimed to accomplish this goal by subsequently testing two hypotheses:

Hypothesis 1. Confidence in CPS will be linked to CPS (i.e., knowledge acquisition and knowledge application as well as CPS variance that is shared by the two facets).

Hypothesis 2. Confidence in CPS will be linked to CPS when reasoning is controlled for (i.e., knowledge acquisition and knowledge application as well as CPS variance that is shared by the two facets).

2. Method

2.1. Participants and procedure³

The sample consisted of 505 seventh-grade students ($M_{age} = 13.16$, $SD_{age} = 0.50$, 48.9% female; 8.4% did not report gender) in the federal German states of Saarland and Rhineland-Palatinate. However, we excluded 38 outliers after computing regressions (e.g., the fourth knowledge acquisition item was regressed on the second and third knowledge acquisition items). If the resulting standardized residuals exceeded two standard deviations, we excluded the outliers, as suggested by Field (2013).⁴ The schools participated voluntarily. The students were enrolled in 21 classes at seven different secondary schools. In Germany, students can acquire one of three different school diplomas: a general diploma, an intermediate diploma, or the highest diploma that enables university admission. The present sample consisted of 13 classes in schools that offered only the highest school diploma (*Gymnasium*), five classes in schools that offered both the general and the intermediate school diplomas (*Realschule Plus*), and two classes in schools where any of the three diplomas could be acquired (*Gesamtschule*). Students were tested in a group setting in their own classes, supervised by two external, trained test administrators. The test battery was comprised of, among others, a paper-and-pencil assessment of fluid reasoning, a CPS test, and measures of confidence in CPS.

2.2. Missing data

The percentage of missing values in the variables ranged from 2.9% to 21.5%. More precisely, the percentage of missing data for the reasoning indicators was 2.9%, 2.9%, 3.1%, and 3.5% in the first test half and 3.1%, 2.9%, 3.1%, and 4.2% in the second test half. The missing rate for knowledge acquisition was 4.0%, 4.4%, 4.4%, and 5.6%. The missing rate for knowledge application was 4.0%, 4.2%, 4.4%, and 5.6%. The missing rate for confidence in knowledge acquisition was 14.2%, 14.2%, 14.4%, and 18.0%, and it was 18.8%, 18.0%, 18.0%, and 21.5% for confidence in knowledge application. Missing data were mainly the result of technical problems, exclusion of the questionnaire due to time restrictions, and a very small proportion of students who did not attend school because of sickness or doctors' appointments. Naturally, data were missing at the highest rates for the

² A priori, we also aimed to investigate *problem solving self-concept* (PSSC) in the nexus of CPS, confidence in CPS, and reasoning using the German adaption (Schwanzer, Trautwein, Lüdtke, & Sydow, 2005) of the Self-Description Questionnaire for young adults (SDQ III; Marsh, 1992). However, due to reliability issues, we decided to exclude PSSC from the analyses in the present manuscript.

³ Please note that the data we are reporting here are part of a larger collection of data and have been used in other manuscripts, too. For instance, Krkovic, Wüstenberg, and Greiff (2016) investigated the structure of collaborative problem solving and CPS.

⁴ This did not change the results notably but led to better scale reliability.

confidence ratings, as these were included in the CPS assessment, which sometimes faced technical problems. Thus, students with missing data on the CPS items were likely to have missing data on the respective confidence ratings. Thus, the missing values should be missing at random, which was also indicated by Little's (1988) test, which showed that the missing values did not occur completely at random ($\chi^2 = 2114.439$, $df = 1927$, $p = .002$).

To impute the missing data, we used multiple imputation that was based on a variance-covariance model (Asparouhov & Muthén, 2010). To impute the missing values of all variables included in the analysis, we used Mplus version 7.11 (Muthén & Muthén, 1998–2014) to generate five complete data sets.⁵

2.3. Measures

2.3.1. CPS (knowledge acquisition and knowledge application)

CPS was assessed with the computer-based MicroDYN approach, which was implemented and conducted with the CBA Item-Builder (Rölke, 2012). Each of the eight MicroDYN items consisted of varying numbers of inputs (e.g., different training methods to train a handball team displayed as Trainings A, B, and C in Fig. 1) and outputs (e.g., motivation, power of throw, and exhaustion in handball). The input variables could be manipulated by moving the slide controls to the left or right and by clicking “apply.” When the slide controls were moved to the right, the inputs (e.g., Training A) increased. If an output (e.g., motivation) was influenced by the input, it increased too. An increase in the output could be seen in an increase in the black graph as it moved upwards and the corresponding increase in the number underneath the output label. An input could be linked to zero, one, or multiple outputs.

First, in the knowledge acquisition phase, the students were asked to explore the task-specific system characteristics (i.e., the relations between inputs and outputs) by changing the slide controls manually. In the example displayed in Fig. 1, the students had to explore which of the training methods had an influence on the output variables “Motivation”, “Power of throw”, and “Exhaustion”. The knowledge about the influences of the inputs that was gained via exploration and was assessed by the task of drawing a graphic model was displayed at the bottom of the screen (Fig. 1; the blue arrows in the model depict an influence of Training A on “Motivation”, an influence of Trainings B and C on “Power of throw”, and an influence of Training C on “Exhaustion”). When students finished drawing the graphic model, they had to press the button labeled “done” and could then move on to the next phase, the knowledge application phase.

In the knowledge application phase, the students were asked to reach a predefined goal state (i.e., the state of the outputs) that was marked with red lines in the outputs and with numbers to the right of the outputs (Fig. 1). To ensure that the students did not use the knowledge application phase to explore the relations between the inputs and outputs (which was what they were supposed to do in the knowledge acquisition phase and which had already been scored with the graphic model), they could click on the “apply” button only a maximum of four times (knowledge application phase; Greiff, 2012). As knowledge acquisition and knowledge application are considered to be two psychometrically separable facets of CPS (Funke, 2001), the students were shown the correct underlying system structure during

⁵ The variables used in the imputation were selected on the basis of theoretical considerations (i.e., we chose the variables that were potentially conceptually linked to the imputed variables) and were comprised of mean scores of cognitive measures and noncognitive measures. However, after including all variables, the imputation model had been too comprehensive to reach convergence. Thus, some variables had to be excluded. The following variables and scales (means) were used in the final imputation model (besides the items from the analyses): grade in mathematics, grade point average, time on task in knowledge acquisition, time on task in knowledge application, need for cognition, all Big Five personality factors, academic self-concept, self-concept of intellect, self-concept in honesty, ICT usage, ICT skills, motivation to work on a computer, numerical reasoning, spatial thinking, effort in working on the CPS tasks, and all PSSC items.

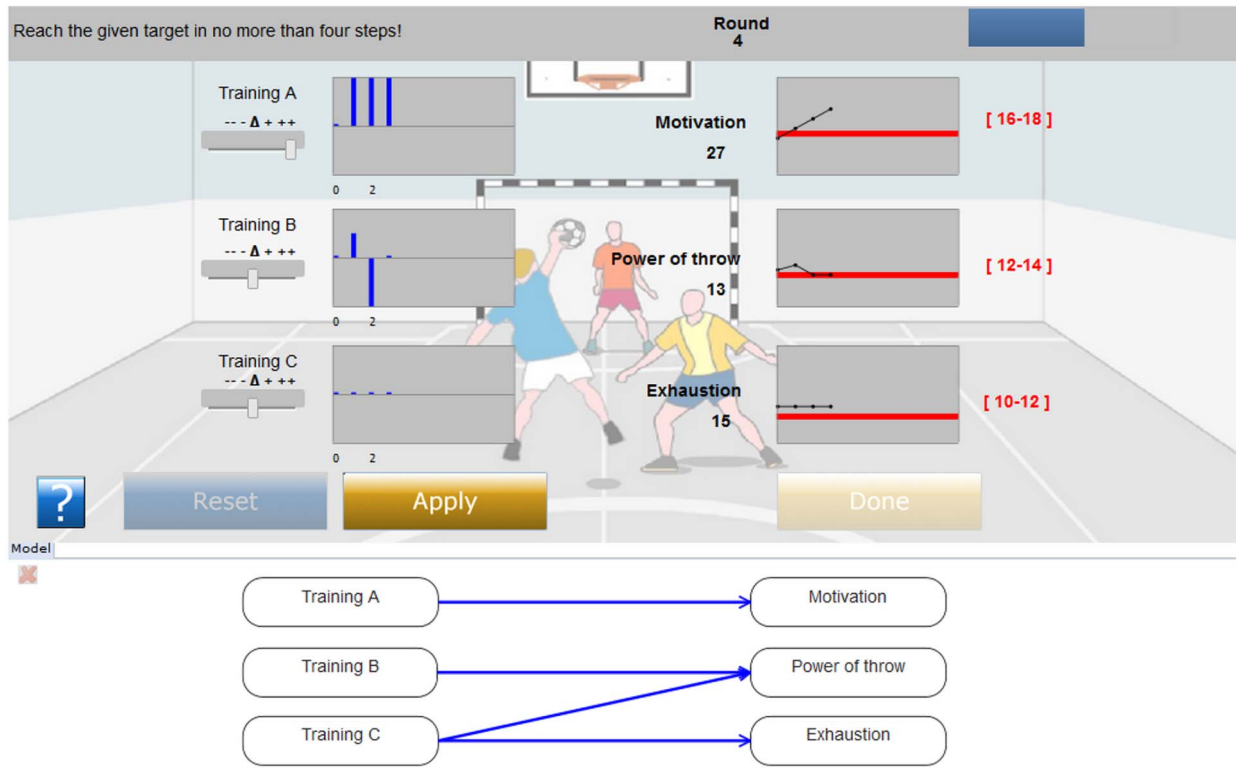


Fig. 1. A translated example of a MicroDYN item (the original item was presented in German) during the knowledge application phase. The red horizontal lines indicate the goal states that need to be reached by manipulating the input variables. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the knowledge application phase, even if they did not arrive at the correct solution during knowledge acquisition (Greiff, 2012). In this way, the students who solved the knowledge acquisition task incorrectly were given the chance to solve the respective knowledge application phase correctly.

For each task, the two phases (i.e., knowledge acquisition and knowledge application) were scored dichotomously. In the knowledge acquisition phase, credit was given for a correctly drawn model, whereas in the knowledge application phase, credit was given if the goal state was reached. As we were interested in the extent to which confidence and reasoning provide explanatory value for shared and unique variance in the CPS facets, we modeled CPS as two separate but correlated factors to represent each facet (knowledge acquisition and knowledge application).

For our study, the instructional phase for MicroDYN, which consisted of three consecutive instructional videos, was supplemented by an additional video that explained the confidence assessment. However, a pilot study with $N = 93$ students showed that the students were overwhelmed by these expanded instructions. This became apparent when the students seemed to be surprised when the pop-up windows for assessing their confidence appeared. Thus, we decided to provide instructions for only the CPS test first and excluded the confidence assessment from the first four items. After administering four MicroDYN items, we gave the instructions for the confidence assessment. In the subsequent four MicroDYN tasks, we assessed students' task-related confidence in addition to their CPS achievement. Thus, we administered a total of 16 indicators for CPS achievement (8 for knowledge acquisition and 8 for knowledge application) and 8 indicators for confidence in CPS (4 for knowledge acquisition and 4 for knowledge application). However, we used only the items that also included the confidence assessments as indicators of students' CPS as our goal was to compare students' achievement directly with their task-related confidence, and we had no measure of students' confidence in their answers to the first four items.

2.3.2. Confidence in CPS (knowledge acquisition and knowledge application)

After the first four MicroDYN items were completed, the students watched a video that explained the additional confidence assessment that was included in the remaining four MicroDYN items. Confidence in CPS was assessed once in each knowledge acquisition phase and once in each knowledge application phase. Confidence in knowledge acquisition was assessed at the end of the knowledge acquisition phase when the students had drawn the model representing the system structures. They were asked if they believed that they had drawn the graphical model correctly or not, which resulted in binary variables. In the subsequent knowledge application phase, we could not assess their confidence after the task (as was done in the knowledge acquisition phase) because the students were able to actually see whether they had reached the goal state or not. Thus, the confidence scores were collected after the first step toward the goal state (i.e., after they clicked on the “apply” button for the first time in the knowledge acquisition phase). The students were able to see the consequences of their first step toward the goal state. Then, the students were asked whether or not they believed they were on the right track toward reaching the goal state. Thus, the confidence assessment resulted in a total of four items for confidence in CPS in the knowledge acquisition phase and four items for confidence in CPS in the knowledge application phase.⁶

2.3.3. Reasoning

Fluid reasoning was assessed with the German version of the Culture

⁶Note that the students were subsequently asked how sure they were about their performance judgment (e.g., “I think I drew the model correctly/incorrectly” followed by “I am very uncertain/.../very certain about this answer”). However, this paper focused on performance judgments. The inclusion of the certainty measure would not have led to different results or conclusions. However, it would have been accompanied by a series of problems (e.g., Is a student who is “very uncertain” that he/she solved a problem “correctly” more confident than a student who is “very uncertain” that he/she solved a problem “incorrectly?”). To ensure the use of sound data, we used only the dichotomous performance judgment score.

Fair Test (CFT 20-R; [Weiß, 2008](#)), which built on the work of [Cattell and Cattell \(1963\)](#). The CFT 20-R consisted of two halves. Each of the two halves was further comprised of four subtests, containing different types of tasks (series, classifications, matrices, and typologies). For each subtest, detailed standardized instructions were given in written form as well as verbal and blackboard-based lessons taught by trained administrators.

The test resulted in a total of 101 dichotomous items. To ensure a parsimonious measurement model, we used four indicators, each of which represented the mean score on each respective subtest (series, classifications, matrices, and typologies). As the number of items on each subtest varied between the test halves, we first computed the mean of each subtest in the first and second test half separately. Subsequently, we built the final indicators for each subtest as a mean of the means of each subtest in each test half. For instance, to compute the indicator for the subtest “series,” we computed the mean score of the series subtest in the first half of the CFT and the mean score of the series subtest in the second half. The score used in the reasoning measurement model was the mean of these two means.

2.4. Statistical analyses

We examined the measurement models for each construct with separate confirmatory factor analyses. We used Mplus 7.11 ([Muthén & Muthén, 1998–2014](#)) for all analyses. The confirmatory factor analysis for reasoning that included continuous data was conducted with Maximum Likelihood (ML) estimation. As the items assessing CPS and confidence were dichotomous, we investigated their factor structure using the weighted least squares means and variance adjusted estimator (WLSMV), which is appropriate for ordinal data ([Muthén & Muthén, 2010](#)).

Even though we did not hypothesize a particular causal relation between confidence and the CPS facets, we chose to conduct multiple latent regression analyses in which the CPS facets were regressed on the confidence facets (and reasoning). We did this to reflect the paradigm of CPS research, to obtain effect sizes (i.e., R^2), and to investigate the extent to which the residuals of the CPS facets remained correlated (i.e., the extent to which the CPS facets remained correlated when confidence in CPS and reasoning were controlled for).

We took the hierarchical structure (students nested in classes) into account by adjusting the standard errors (with the “type is complex” option implemented in Mplus) in all analyses, and we used the CFI, TLI, and RMSEA as indicators of the goodness of fit for all factor analyses and regressions. As we imputed five data sets, all reported fit criteria are mean values from all five imputed data sets.⁷

To estimate the reliability of the constructs, we used McDonald's Omega ([McDonald, 1970](#)) for all scales with metric indicators. For the constructs that were assessed with binary data, we used [Green and Yang's \(2009\)](#) indicator, which is appropriate for categorical data. [Goldhammer, Kröhne, Keßel, Senkbeil and Ihme \(2014\)](#) adapted [Green and Yang's \(2009\)](#) SAS code for R. This script was also used in the present study.

3. Results

3.1. CPS measurement models (knowledge acquisition and knowledge application)

We used four indicators for knowledge acquisition and four indicators for knowledge application, respectively. As we were interested in the extent to which the hypothesized predictors could explain

⁷ The confidence intervals for the RMSEA values were not available because [Browne and Cudeck's \(1993\)](#) method has not been generalized to analyses with multiple imputations and the WLSMV estimator.

the variance that was shared by knowledge acquisition and knowledge application, we did not model CPS as a hierarchical model but as two separate constructs that are correlated.

The models for the CPS factors showed good fit (knowledge acquisition: CFI = 0.998; TLI = .993; RMSEA = .055; knowledge application: CFI = .998; TLI = .993; RMSEA = .039) and had sufficient reliability: .83 for knowledge acquisition and .70 for knowledge application. The high model fit values can be accepted with some caution because the number of parameters in the confirmatory factor analyses was only slightly higher than the number of cells in the covariance matrices that were the basis for the model estimation (this holds for all factors in this paper). However, the models that we computed to address the hypotheses also contained the measurement models and indicated a good overall fit. Thus, we could conclude that the measurement models represented the given data well.

The mean of all manifest knowledge acquisition items was higher ($M = .46$, $SD = .40$) than the mean of all knowledge application items ($M = .31$, $SD = .30$), thus indicating that knowledge acquisition was slightly easier for the students than knowledge application.

3.2. Confidence in CPS measurement models (knowledge acquisition and knowledge application)

The models for confidence showed very good fits for both CPS factors (knowledge acquisition: CFI = .998; TLI = .994; RMSEA = .038; knowledge application: CFI = .999; TLI = .999; RMSEA = .017) and good reliability (knowledge acquisition: .79; knowledge application: .80). The mean of the manifest confidence items showed that, overall, the students' confidence was much higher than their actual performance (mean confidence for knowledge acquisition: $M = .77$, $SD = .33$; mean confidence for knowledge application: $M = .81$, $SD = .31$).

3.3. Reasoning measurement model

A latent reasoning factor model with four indicators (means of series, classifications, matrices, and typologies for a task) showed a satisfying model fit (CFI = .995; TLI = .985; RMSEA = .063) and reliability ($\omega = .85$). The mean was $M = .64$ ($SD = .13$). As expected, the reliability of the measure was higher when both test halves were used than when only one half was used ($\omega_1 = .77$; $\omega_2 = .75$), which confirms that both test halves of the CFT should be used as an indicator for reasoning.

3.4. Correlations between constructs

[Table 1](#) displays the manifest correlations (between the sum scores) and the latent correlations (between the factor scores). The manifest correlations were computed in SPSS as correlations between the mean values of all of the items used to measure a construct. The correlations of the latent factors were computed in Mplus 7.11 in a model in which all of the latent constructs were correlated (CFI = .991; TLI = .989; RMSEA = .021). The correlations showed that the proposed latent predictors of CPS (confidence in CPS and reasoning) were significantly ($p < .001$) and strongly ($\rho \geq .63$; $r \geq .37$) correlated with both CPS facets.

3.5. Preliminary analyses

To further verify that reasoning is a substantial predictor of CPS and to later be able to quantify the proposed incremental variance accounted for by confidence in explaining CPS, we estimated a model in which only reasoning predicted the CPS facets. This model fit well (CFI = .995; TLI = .994; RMSEA = .019) and explained 52% of the variance in knowledge acquisition ($p < .001$) and 55% of the variance in knowledge application ($p < .001$).

Table 1
Correlation matrix for all factors in the analyses.

	CPS ACQ	CPS APP	Confidence in ACQ	Confidence in APP	Reasoning
CPS ACQ	–	.75***	.85***	.63***	.72***
CPS APP	.53***	–	.63***	.72***	.74***
Confidence in ACQ	.56***	.38***	–	.84***	.54***
Confidence in APP	.39***	.37***	.60***	–	.55***
Reasoning	.56***	.51***	.41***	.42***	–

Note. Correlations between factor scores (as represented by the latent variables) above the diagonal; correlations between mean scores of all indicators of each factor below the diagonal; CPS = complex problem solving; ACQ = knowledge acquisition; APP = knowledge application. *** $p < 0.001$.

3.5.1. Hypothesis 1 confidence in CPS will be linked to CPS

To understand the link between confidence in CPS and CPS performance, we regressed the two CPS dimensions, knowledge acquisition and knowledge application, on confidence in CPS in knowledge acquisition and confidence in CPS in knowledge application, respectively (Fig. 2). More precisely, we regressed CPS knowledge acquisition on confidence in knowledge acquisition and CPS knowledge application on confidence in knowledge application. The crosswise paths (i.e., regressing CPS knowledge acquisition on confidence in knowledge application and CPS knowledge application on confidence in knowledge acquisition) were fixed to zero.

The model fit the data well (CFI = .990; TLI = .988; RMSEA = .027) and suggested that confidence in CPS was substantially linked to CPS, as the respective confidence factor explained 67% of the variance in knowledge acquisition ($p < .001$) and 55% of the variance in knowledge application ($p < .001$). Further, we found that the residuals for knowledge acquisition and knowledge application remained correlated at $\rho = .67$ ($p < .001$). In spite of the excellent model fit, one might have been interested in including the crosswise paths (i.e., regressing CPS knowledge acquisition on confidence in knowledge application and CPS knowledge application on confidence in knowledge acquisition). This, however, would have led to estimation problems as one of the standardized regression weights was greater than one.

3.5.2. Hypothesis 2 confidence in CPS will be linked to CPS beyond reasoning

To investigate the extent to which confidence in CPS is linked to CPS beyond reasoning, we estimated a similar structural equation model as

depicted in Fig. 3. This time we included reasoning as an additional predictor of the CPS facets knowledge acquisition and knowledge application (Fig. 3).

This model fit the data well (CFI = .989; TLI = 0.987; RMSEA = .023) and showed that the link between the CPS factors and the respective confidence in the CPS factors remained substantial even when reasoning was controlled for (knowledge acquisition: $\beta = .61$, $SE = .06$, $p < .001$; knowledge application: $\beta = .45$, $SE = .09$, $p < .001$).

Moreover, the model showed the extent to which reasoning and confidence in the respective CPS facet together explained each CPS facet. This result came to 76% of the variance in knowledge acquisition ($p < .001$) and 69% of the variance in knowledge application ($p < .001$). To obtain information on the incremental variance that confidence predicted in CPS beyond reasoning, we compared the model as displayed in Fig. 3 with a model in which spite of the excellent model fit, one only reasoning (but not the confidence factor) predicted the CPS facet (see Preliminary Analyses). The preliminary model without confidence in CPS explained 52% of the variance in knowledge acquisition ($p < .001$) and 55% of the variance in knowledge application ($p < .001$). Thus, confidence explained an increment of 24 percentage points of the variance in knowledge acquisition and 14 percentage points of the variance in knowledge application.

The results in Fig. 3 further show that the correlation between the residuals of the CPS facets was reduced substantially ($\rho = .34$, $SE = .15$, $p = .022$). This indicates that the variance that is shared by the CPS facets can be explained to a great extent by reasoning and confidence in CPS.

4. Discussion

The aim of the present study was to advance our understanding of the nature of CPS by investigating the link between metacognitive monitoring and CPS performance. Confidence in CPS as an indicator of metacognitive monitoring was expected to be substantially linked to CPS because successful CPS requires students to continuously monitor their problem-solving behavior to be able to adapt it when necessary. The decision about whether adaption is necessary relies on students' confidence. To ensure that the link between confidence in CPS and CPS is not an artefact of the known relation between these variables and reasoning skills, we controlled for reasoning. The results indicated that confidence in CPS was still strongly linked to CPS performance even after reasoning had been controlled for. The results will be discussed in more detail below.

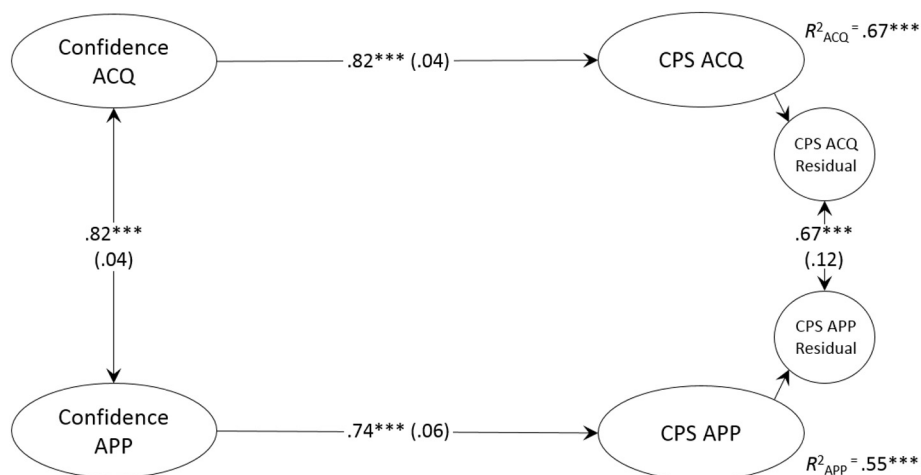


Fig. 2. CPS = complex problem solving; ACQ = knowledge acquisition; APP = knowledge application; standard errors in parentheses. *** $p < .001$.

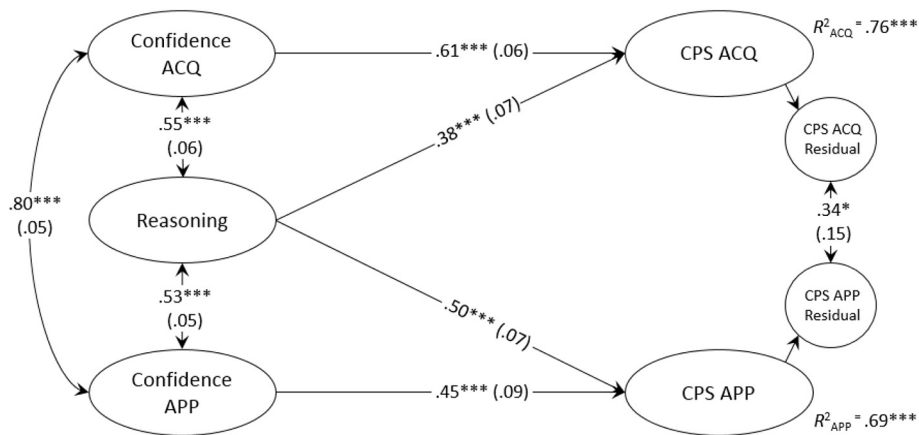


Fig. 3. CPS = complex problem solving; ACQ = knowledge acquisition; APP = knowledge application; standard errors in parentheses. * $p < .05$. *** $p < .001$.

4.1. Confidence in CPS is linked to CPS

The substantial link between confidence in CPS and CPS indicates that confidence in CPS does have substantial value in explaining CPS, thus validating the proposal that successful CPS requires efficient self-regulation, particularly metacognitive monitoring. The reason for this might be that students need to realize when they have to change their problem-solving behavior to be successful self-regulated problem solvers (Funke, 2003; Greiff & Fischer, 2013; Ifenthaler, 2012; Mayer, 1998). Confidence judgments might provide an important basis for such metacognitive activities. Moreover, our results indicate that students are mostly aware of their performance in CPS and should, in general, be able to utilize task-inherent cues as presented in CPS tasks (Kröner & Biermann, 2007).

Our results show that confidence and performance scores are more closely related during knowledge acquisition than during knowledge application. This may be explained by the task characteristics: In the knowledge acquisition phase, there was plenty of time to adapt one's problem solving behavior, whereas the time and number of clicks were limited in the knowledge application phase such that self-regulation might have a more limited effect in the knowledge application phase.

4.2. Confidence in CPS is linked to CPS beyond reasoning

As mentioned above, the link between confidence in CPS and CPS performance remained stable when we controlled for reasoning, suggesting that self-monitoring is an essential component of CPS. Future studies may wish to use this as a starting point to investigate discriminant as well as differential predictive validity of performance judgments on various tests. For this goal, confidence should be included not only in CPS but also in other performance measures applied in future studies. This accounts in particular for confidence in reasoning, as reasoning and CPS are closely related (Stadler et al., 2015).

The results further show that confidence in CPS and reasoning explain a substantial amount of the CPS-specific variance that is shared by the two CPS facets knowledge acquisition and knowledge application. This indicates that self-monitoring and reasoning are core components that are necessary for making a person a good complex problem solver. An interesting open question that remains is whether other, more distal constructs than those investigated here may demonstrate an interplay with performance judgments to explain performance in CPS (e.g., mastery orientation; Scherer, Greiff, & Hautamäki, 2015).

4.2.1. Limitations

When interpreting the presented results, some limitations need to be taken into account. For instance, this study did not allow any causal inferences to be made due to its cross-sectional nature. In addition, the study was based on the voluntary participation of the schools and

students, who also needed the consent of their parents. Thus, the representativeness of our sample for the population of German students cannot be taken for granted.

Another limitation involves the extent to which confidence was assessed: Confidence consists of two aspects: the judgment (i.e., whether the students think they solved a task correctly or not) and the extent to which they are certain about their performance judgment. As mentioned in the beginning, the wording of the confidence assessment in our study confounds the two aspects, as in most existing studies. We focused on the performance judgment as a central aspect of CPS. Future studies may wish to build on the present findings and investigate the relevance of the certainty of the performance judgment while disentangling performance and certainty judgments in the wording of the question (Händel & Fritzsche, 2016). However, we expect both aspects to be valid indicators of self-monitoring in CPS tasks, as problem solvers are in constant need of evaluating their performance during problem solving to decide whether they should continue following the path and strategy they chose or whether they should change their problem solving approach (Funke, 2003; Mayer, 1998).

4.2.2. Outlook

The present study contributes to a better understanding of CPS as it suggests that self-regulation, particularly metacognitive monitoring as reflected by confidence in CPS, is closely linked to CPS. This finding is in line with previous findings that have shown that CPS is linked to other aspects of metacognition. In particular, Wüstenberg et al. (2014) postulated that metastrategic knowledge (i.e., knowledge about knowing a strategy) is relevant for successful CPS. Future research may wish to build on this foundation and investigate additional aspects of metacognition and their role in CPS. It might be especially interesting to investigate the adaption of the CPS process after a negative confidence judgment, as we believe that this is at the core of CPS: Students do not profit from simply knowing that their problem solving is improvable, but they need to know how to improve it, too.

Funding

This research was funded by the Fonds National de la Recherche Luxembourg (ATTRACT "ASKI21"; AFR "SINQ"; CORE "DynASCEL").

References

Asparouhov, T., & Muthén, B. (2010). *Multiple imputation with Mplus*. (Technical Report, Version 2).
 Babcock, R. L. (2002). Analysis of age differences in types of errors on the Raven's advanced progressive matrices. *Intelligence*, 30, 485–503. [http://dx.doi.org/10.1016/S0160-2896\(02\)00124-1](http://dx.doi.org/10.1016/S0160-2896(02)00124-1).
 Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A.

- Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage. <http://dx.doi.org/10.1177/0049124192021002005>.
- Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch, & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 78–107). New York: Psychology Press.
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5–21). Amsterdam, the Netherlands: Pergamon. <http://dx.doi.org/10.1016/B978-008043793-4/50036-2>.
- Cattell, R. B., & Cattell, A. K. S. (1963). *Culture fair intelligence test*. Champaign, IL: Institute for Personality and Ability Testing.
- Field, A. P. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock 'n' roll* (4th ed.). London: Sage publications.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: a new area of cognitive-developmental inquiry. *American Psychologist*, 34(10), 906–911. <http://dx.doi.org/10.1037/0003-066x.34.10.906>.
- Fritzche, E. S., Kröner, S., Dresel, M., Kopp, B., & Martschinke, S. (2012). Confidence scores as measures of metacognitive monitoring in primary students? (Limited) validity in predicting academic achievement and the mediating role of self-concept. *Journal for Educational Research Online*, 2, 120–142.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgment. *Thinking and Reasoning*, 7, 69–89. <http://dx.doi.org/10.1080/13546780042000046>.
- Funke, J. (2003). *Problemlösendes Denken [Problem solving thinking]*. Stuttgart: Kohlhammer.
- Goldhammer, F., Kröhne, U., Keßel, Y., Senkbeil, M., & Ihme, J. M. (2014). Multiple-Choice- vs. simulationsbasierte Aufgaben. *Diagnostica*, 60, 10–21. <http://dx.doi.org/10.1026/0012-1924/a000113>.
- Greiff, S. (2012). *Individualdiagnostik komplexer Problemlösefähigkeit [assessment of complex problem solving ability]*. Münster: Waxmann.
- Greiff, S., & Fischer, A. (2013). Measuring complex problem solving: an educational application of psychological theories. *Journal for Educational Research Online*, 5, 34–53.
- Greiff, S., Kretschmar, A., Müller, J. C., Spinath, B., & Martin, R. (2014). The computer-based assessment of complex problem solving and how it is influenced by students' information and communication technology literacy. *Journal of Educational Psychology*, 105, 666–680. <http://dx.doi.org/10.1037/a0035426>.
- Greiff, S., Wüstenberg, S., Molnar, G., Fischer, A., Funke, J., & Csapo, B. (2013). Complex problem solving in educational settings – something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, 105, 364–379. <http://dx.doi.org/10.1037/a0031856>.
- Green, S. B., & Yang, Y. (2009). Reliability of summed item scores using structural equation modeling: An alternative to coefficient alpha. *Psychometrika*, 74, 155–167. <http://dx.doi.org/10.1007/S11336-008-9099-3>.
- Händel, M., & Fritzche, E. S. (2016). Unskilled but subjectively aware: Metacognitive monitoring ability and respective awareness in low-performing students. *Memory and Cognition*, 44, 229–241. <http://dx.doi.org/10.3758/s13421-015-0552-0>.
- Huber, O. (1995). Complex problem solving as multistage decision making. In P. Frensch, & J. Funke (Eds.), *Complex problem solving: The European perspective* (pp. 151–173). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.
- Ifenthaler, D. (2012). Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society*, 15, 38–52.
- Kleitman, S., & Gibson, J. (2011). Metacognitive beliefs, self-confidence and primary learning environment of sixth grade students. *Learning and Individual Differences*, 21, 728–734. <http://dx.doi.org/10.1016/j.lindif.2011.08.003>.
- Kleitman, S., & Moscrop, T. (2010). Self-confidence and academic achievements in primary-school children: Their relationships and links to parental bonds, intelligence, age, and gender. In A. Efklides, & P. Misailidis (Eds.), *Trends and prospects in metacognition research* (pp. 293–326). US: Springer. http://dx.doi.org/10.1007/978-1-4419-6546-2_14.
- Kretschmar, A., Neubert, J. C., Wüstenberg, S., & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence*, 54, 55–69. <http://dx.doi.org/10.1016/j.intell.2015.11.004>.
- Krkovic, K., Wüstenberg, S., & Greiff, S. (2016). Assessing collaborative behaviour in students. An experiment-based approach. *European Journal of Psychological Assessment*, 32(1), 52–60. <http://dx.doi.org/10.1027/1015-5759/a000329>.
- Kröner, S., & Biermann, A. (2007). The relationship between confidence and self-concept: Towards a model of response confidence. *Intelligence*, 35, 580–590. <http://dx.doi.org/10.1016/j.intell.2006.09.009>.
- Kröner, S. (2014). Pädagogisch-Psychologische Problemlöseforschung als Problem-Barrieren überwinden. *Zeitschrift für Pädagogische Psychologie*, 28(4), 229–235. <http://dx.doi.org/10.1024/1010-0652/a000139>.
- Kröner, S., Plass, J. L., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33, 347–368. <http://dx.doi.org/10.1016/j.intell.2005.03.002>.
- Kuhn, D. (2000). Metacognitive Development. *Current Directions in Psychological Science*, 9(5), 178–181. <http://dx.doi.org/10.1016/j.lindif.2011.08.003>.
- Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202. <http://dx.doi.org/10.1080/01621459.1988.10478722>.
- Marsh, H. W. (1992). *Self description questionnaire (SDQ) III: A theoretical and empirical basis for the measurement of multiple dimensions of late adolescent self-concept: An interim test manual and a research monograph*. Macarthur, New South Wales, Australia: University of Western Sydney, Faculty of Education.
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science*, 26, 49–63. <http://dx.doi.org/10.1023/a:1003088013286>.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical Psychology*, 23, 1–21. <http://dx.doi.org/10.1111/j.2044-8317.1970.tb00432.x>.
- Muthén, L. K., & Muthén, B. O. (1998–2014). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Rölke, H. (2012). The ItemBuilder: A graphical authoring system for complex item development. In T. Bastiaens, & G. Marks (Eds.), *Proceedings of world conference on e-learning in corporate, government, healthcare, and higher education* (pp. 344–353). Chesapeake, VA: AACE. Retrieved from <http://www.editlib.org/p/41614>.
- Scherer, R., Greiff, S., & Hautamäki, J. (2015). Exploring the relation between time on task and ability in complex problem solving. *Intelligence*, 48, 37–50. <http://dx.doi.org/10.1016/j.intell.2014.10.003>.
- Schraw, G., Crippen, K. J., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education*, 36, 111–139. <http://dx.doi.org/10.1007/s11165-005-3917-8>.
- Schwanzler, A. D., Trautwein, U., Lüdtke, O., & Sydow, H. (2005). Entwicklung eines instruments zur Erfassung des Selbstkonzepts junger Erwachsener [Development of the instruments to assess the self-concept of young adults]. *Diagnostica*, 51(4), 183–194. <http://dx.doi.org/10.1026/0012-1924.51.4.183>.
- Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence. A meta-analysis. *Intelligence*, 53, 92–101. <http://dx.doi.org/10.1016/j.intell.2015.09.005>.
- Stankov, L. (2000). Complexity, metacognition, and fluid intelligence. *Intelligence*, 23, 121–143. [http://dx.doi.org/10.1016/S0160-2896\(99\)00033-1](http://dx.doi.org/10.1016/S0160-2896(99)00033-1).
- Stankov, L., Lee, J., Luo, W., & Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learning and Individual Differences*, 22, 747–758. <http://dx.doi.org/10.1016/j.lindif.2012.05.013>.
- Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100, 961–976. <http://dx.doi.org/10.1037/a0012546>.
- Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education*, 14, 541–554. <http://dx.doi.org/10.1007/bf03172978>.
- Weiß, R. H. (2008). *Grundintelligenztest Skala 2 - Revision - (CFT 20-R)*. Göttingen: Hogrefe.
- Wittmann, W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, traits, and content determinants* (pp. 77–108). Washington, DC: APA.
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving: More than reasoning? *Intelligence*, 40, 1–14. <http://dx.doi.org/10.1016/j.intell.2011.11.003>.
- Wüstenberg, S., Stadler, M., Hautamäki, H., & Greiff, S. (2014). The role of strategy knowledge for the application of strategies in complex problem solving tasks. *Technology, Knowledge and Learning*, 19, 127–146. <http://dx.doi.org/10.1007/s10758-014-9222-8>.
- Zimmermann, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 233–262). Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511615771.009>.