Journal of Educational Psychology
2014, Vol. 106, No. 4, 1007–1020

0022-0663/14/$12.00 http://dx.doi.org/10.1037/a0037114

The Development of Complex Problem Solving in Adolescence: A Latent Growth Curve Analysis

Gidon T. Frischkorn
University of Heidelberg

Samuel Greiff and Sascha Wüstenberg
University of Luxembourg

Complex problem solving (CPS) as a cross-curricular competence has recently attracted more attention in educational psychology as indicated by its implementation in international educational large-scale assessments such as the Programme for International Student Assessment. However, research on the development of CPS is scarce, and the few existing studies are cross-sectional. Therefore, the present study analyzed CPS development with longitudinal data on adolescent students collected over a period of 2 years. CPS development was estimated with latent growth curve models, and fluid reasoning, age, and sex served as predictors. CPS growth patterns were positive and linear and were positively related to fluid reasoning. Older students performed better on initial CPS but exhibited smaller increases (i.e., less development) in CPS performance. No meaningful sex differences in initial CPS or CPS development were found except that boys showed slightly better initial performance on the CPS dimension knowledge application. These results present an important first step in the investigation of CPS development.

Keywords: cognitive development, complex problem solving, MicroDYN, fluid reasoning, latent growth curve analysis

Cognitive abilities such as fluid reasoning are highly relevant in everyday life, especially for adapting to new situations and finding solutions to everyday problems (Mayer & Wittrock, 2006). Many tasks in our lives such as learning new skills or solving problems rely on multiple sets of cognitive abilities (Sackett, Bornemann, & Conelly, 2008; Schmidt & Hunter, 1998). Furthermore, the challenges in working life are no longer limited to routine tasks and manual skills that do not require interaction or problem solving. Nowadays, nonroutine tasks requiring analytical and interactive skills such as collecting and organizing information, problem solving, or finding solutions as a team have become far more important than the aforesaid routine tasks or manual skills (Autor, Levy, & Murnane, 2003).

Recent studies that have investigated cognitive abilities and their importance for success in school and working life have often used complex problem solving (CPS) tasks to assess analytical and interactive skills (e.g., Danner, Hagemann, Schankin, Hager, & Funke, 2011; Sonnleitner, Keller, Martin, & Brunner, 2013; Wüstenberg, Greiff, & Funke, 2012). Conceptually, CPS is about effectively handling new situations, overcoming difficulties by generating new solutions, and mentally representing complex information (Funke, 2010). CPS is therefore considered to be a transversal skill because the skills needed to gain knowledge, acquire information, and build viable representations of everyday problems are not restricted to specific domains (Greiff, Wüstenberg, & Funke, 2012). Transversal skills are often described as cross-curricular skills in education because they are needed across different school subjects.

CPS and other transversal skills have recently received considerable attention in educational contexts. For instance, CPS has been assessed in educational large-scale assessments of school performance (Organization for Economic Co-Operation and Development [OECD], 2004)[1] such as the Programme for International Student Assessment (PISA). In PISA 2012, CPS was selected as an additional domain that complemented the domains of mathematics, science, and reading (OECD, 2010). In addition to measuring performance at a single point in time, measuring the development of these skills is also important in such assessments as well as in educational psychology in general. The development of transversal skills is of particular interest to the field of educational psychology as these skills may enhance individual learning skills across different subjects (OECD, 2004). Mayer and Wittrock (2006) further emphasized that helping students to become better

---

Gidon T. Frischkorn, Department of Psychology, University of Heidelberg; Samuel Greiff and Sascha Wüstenberg, Education, Culture, Cognition and Society Unit, University of Luxembourg.

This research was funded by grants supported by the Fonds National de la Recherche Luxembourg (ATTRACT “ASSKI21”) and by the European Union (290683; LLLight in Europe). We are grateful to the TBA group at DIPF (http://tba.dipf.de) for providing the authoring tool CBA Item Builder and technical support.

Correspondence concerning this article should be addressed to Samuel Greiff, ECCS Unit, University of Luxembourg, 6, rue Richard Coudenhove Kalergi, 1359 Luxembourg-Kirchberg, Luxembourg. E-mail: samuel.greiff@uni.lu

[1] In recent research, different labels for solving dynamic, complex, and interactive tasks such as *complex problem solving* (e.g., Greiff et al., 2012) or *interactive problem solving* (e.g., OECD, 2010) have been used synonymously to emphasize different aspects of the task. In this study, we consistently use the term complex problem solving.

problem solvers is among the most important challenges that educational psychology will have in the 21st century.

However, research on CPS, and longitudinal research in particular, is scarce compared with other domains such as math, science, or literacy (e.g., Berninger & Abbott, 2010; Crosnoe et al., 2010; Vaessen et al., 2010), all of which are well represented in school curricula. Meanwhile, CPS is underrepresented in schools despite its soaring importance in society (Autor et al., 2003) and in educational psychology (Mayer & Wittrock, 2006). The underrepresentation of CPS in school curricula may be one reason for the lack of longitudinal research in the area of CPS. Nevertheless, deeper knowledge about the development of CPS is needed in the field of educational psychology. Such knowledge may be fundamental for further research on training methods or interventions for enhancing CPS skills in today's students. Therefore, we conducted a longitudinal investigation on the development of CPS skills in adolescents to provide some first insights into the development of CPS. Further, we tried to integrate CPS and its development into the structure of cognitive abilities and other correlates of cognitive development. To do so, we additionally investigated how fluid reasoning, age, and sex were related to CPS and its development.

CPS

CPS is considered a complex cognitive skill (Funke, 2010; Klahr, 2000) that is characterized by two core cognitive processes: (a) the need to acquire information about a given opaque problem situation and to represent the information in a viable and parsimonious way (knowledge acquisition) and (b) the use of this knowledge to reach given or self-formulated goals within the problem scenario (knowledge application; Fischer, Greiff, & Funke, 2012; Funke, 2001; Kröner, Plass, & Leutner, 2005). The cognitive processes underlying CPS thus direct a person's actions as he or she interacts with dynamic task environments and integrates initially unknown information in order to achieve certain goals by following consecutive and contingent steps (Buchner, 1995; Raven, 2000). Therefore, CPS is by nature a nonroutine skill that goes beyond routine actions to a considerable degree (Funke, 2010; Greiff, 2012b).

Regarding its relation to other cognitive abilities, CPS may be tentatively integrated into the framework of the Cattell-Horn-Carroll (CHC) theory of intelligence (McGrew, 2009). The CHC theory combines two major theories of intelligence, the Cattell-Horn theory of fluid and crystallized intelligence (Horn & Blankson, 2005; Horn & Noll, 1997) and the three-stratum theory of intelligence by Carroll (2005). It thereby takes Carroll's idea of three strata that contain more or less general cognitive abilities and combines this idea with the separation of different cognitive abilities from the Cattell-Horn theory of intelligence (McGrew, 2009). On the most general stratum, Stratum III, there is only one general cognitive ability, *g*. The second Stratum, Stratum II, includes broad cognitive abilities such as fluid reasoning or short-term memory. These broad cognitive abilities are then divided into narrow cognitive abilities on Stratum I.

Regarding the integration of CPS into the CHC theory, CPS is expected to have some overlap with other broad cognitive abilities located on Stratum II in the CHC theory, such as fluid reasoning (Wüstenberg et al., 2012), because the two rely to a certain extent on similar cognitive processes (Babcock, 2002; Wiley, Jarosz, Cushen, & Colfiesh, 2011). For example, actions such as generating a mental representation of a task or applying abstract rules to solve a task are needed for fluid reasoning tasks (Babcock, 2002) and CPS tasks alike (Wüstenberg et al., 2012). Some of these overlapping cognitive processes may well be viewed as parts of Piagetian reasoning (Hooper, Fitzgerald, & Papalia, 1971). This conceptual relation is reflected in positive correlations between CPS and fluid reasoning measures, ranging from about .40 to .70 (Gonzalez, Thomas, & Vanyukov, 2005; Wenke, Frensch, & Funke, 2005; Wüstenberg et al., 2012).

Nevertheless, there are a number of major conceptual differences between CPS and fluid reasoning (Wüstenberg et al., 2012). For example, CPS requires the problem solver to actively interact with the problem scenario in order to generate knowledge hidden at the outset and to use this knowledge by applying procedural abilities to control the problem scenario (Buchner, 1995; Funke, 2010; Greiff et al., 2012). These characteristics are hardly represented in fluid reasoning (Raven, 2000). Therefore, CPS differs considerably from fluid reasoning and requires cognitive processes other than fluid reasoning (Funke, 2010).

Despite the positive relations between CPS and fluid reasoning, considerable parts of the variance in CPS cannot be empirically explained by fluid reasoning. Previous studies have also reported incremental validity for CPS beyond fluid reasoning and other cognitive abilities in predicting school performance, academic success, or supervisory ratings of overall job performance (Danner, Hagemann, Schankin, et al., 2011; Schweizer, Wüstenberg, & Greiff, 2013; Wüstenberg et al., 2012). In summary, CPS may be tentatively integrated into the CHC theory on the second stratum of broad cognitive abilities (Wüstenberg et al., 2012).

These previous results provide important indications of the relevance of CPS for educational psychology. However, in addition to knowledge about individual differences in CPS and relations between CPS and other cognitive abilities, knowledge about the development of CPS and its facilitation is important because the development of cognitive abilities in educational contexts (e.g., through schooling) is highly relevant for educational achievement and later success in working life (Deary, Strand, Smith, & Fernandes, 2007).

With regard to the development of CPS, recent cross-sectional studies have reported age differences in CPS skills (Greiff, Wüstenberg, et al., 2013; Molnár, Greiff, & Csapó, 2013) such that older students had better CPS performance. These investigations represent the first step in investigating CPS development. Hence, the goal of the present study was to provide the first longitudinal examination of the development of CPS in adolescents in order to specifically address the development of CPS as a complex cognitive skill in school. A further goal was to provide basic knowledge that could be used to launch further research on interventions designed to help students develop their CPS skills.

The Development of Cognitive Abilities

Cognitive abilities progress from birth until the end of late adolescence (Jensen, 2006; Nettelbeck & Burns, 2010). Hence, absolute scores on tasks of cognitive performance tend to increase up to the ages of 20–22, a process that has been reported and extensively discussed theoretically for a number of abilities such as short-term memory and fluid reasoning (e.g., de Jong & de

Jong, 1996; Fry & Hale, 1996; Kail, 2007; Romine & Reynolds, 2005). The time frame and shape of cognitive development are particularly relevant because both may differ notably across various cognitive abilities (Campbell, Pungello, Miller-Johnson, Burchinal, & Ramey, 2001; Kail, 1991). The development of fluid abilities, for example, precedes the development of crystallized abilities and thus takes place at earlier ages (Cattell, 1971). Regarding the development of CPS, Molnár et al. (2013) found in a cross-sectional study that the largest gains in CPS development occurred around the age of 13 years. Thus, CPS skills appear to develop in adolescence rather than in childhood.

The relations between different cognitive abilities and their development are particularly important. Cognitive development has often been described theoretically as the subsequent development of cognitive abilities (Carpenter, Just, & Shell, 1990; Kail, 1992). Specifically, cognitive abilities that develop early may enhance the later development of other cognitive abilities. With regard to this progress in cognitive development, earlier cross-sectional and longitudinal research has reported that the development of fluid reasoning is preceded by growth in short-term memory and faster processing speed, indicating that short-term memory and processing speed enhance the development of fluid reasoning (Coyle, Pillow, Snyder, & Kochunov, 2011; Fry & Hale, 1996; Kail, 2007).

Additionally, recent investigations have reported substantial relations between fluid reasoning in adolescence and executive functioning in early childhood (e.g., Richland & Burchinal, 2013; Thibaut, French, & Vezneva, 2010). Executive functioning is composed of simple cognitive processes and can be measured with different tasks such as the Tower of Hanoi or the Children's Stroop task. It is a regulatory process of cognitive action that develops in early childhood and facilitates the development of cognitive abilities such as planning, switching tasks, or controlling attention (Diamond, 2002; Stuss, 2007). This example further demonstrates that cognitive abilities that develop early, such as executive functioning, may explain the later development of other cognitive abilities such as fluid reasoning. As CPS develops in adolescence rather than in early childhood (Molnár et al., 2013), we suggest that CPS and its development rely on fluid reasoning just as fluid reasoning and its development rely on executive functioning. Thus, CPS may be positioned next in line after fluid reasoning in the developmental cascade suggested earlier (Kail, 1992).

The Present Study

The present study included CPS scores on the two aforementioned CPS dimensions (i.e., knowledge acquisition and knowledge application) at three measurement points to assess the development of CPS in adolescent students. Thereby, we were able to conduct a longitudinal investigation of CPS development by separating *initial CPS performance* at the first measurement point from *CPS development* measured as differences in CPS performance between initial CPS performance and CPS measured at later time points.

We included fluid reasoning as a predictor of initial CPS and CPS development because fluid reasoning is theoretically and empirically related to CPS (Greiff, Fischer, et al., 2013; Wüstenberg et al., 2012) and because cognitive abilities that develop earlier predict the subsequent development of other cognitive abilities (Kail, 2007; Nettelbeck & Burns, 2010). We also included age as a predictor to detect differences in initial CPS and CPS development related to age. We additionally considered sex as a predictor of initial CPS and its development because earlier investigations have reported advantages in CPS for boys compared with girls (Wittmann & Hattrup, 2004; Wüstenberg, Greiff, Molnár, & Funke, 2014). In the following, we outline three research questions (RQs) for the present study.

RQ1: Development of CPS

We expected that CPS skills would increase with time because absolute scores on tests of cognitive ability increase with age until late adolescence (de Jong & de Jong, 1996; Nettelbeck & Burns, 2010). On the basis of Stelzl, Merz, Ehlers, and Remer's (1995) finding that schooling and the maturation of children and adolescents led to rather linear increases in fluid reasoning, we expected that CPS skills would also increase linearly. This development would be reflected in a significant positive linear slope and steady increases in CPS performance between subsequent measurement points. However, a recent study reported that the development of fluid abilities such as processing speed slows down around the ages of 20–22 (Nettelbeck & Burns, 2010). Thus, we also included a quadratic term to test for a negatively accelerated (i.e., quadratic) developmental pattern. Relations between initial CPS and CPS development were additionally evaluated to take a closer look at relations between initial CPS and its development.

RQ2: Fluid Reasoning as a Predictor of CPS Development

We expected that fluid reasoning would positively predict initial CPS performance because CPS is theoretically and empirically related to fluid reasoning (Danner, Hagemann, Schankin, et al., 2011; Schweizer et al., 2013; Wüstenberg et al., 2012). However, this relation was expected to indicate only moderate relations between fluid reasoning and initial CPS skills (Greiff, Wüstenberg, et al., 2013; Wüstenberg et al., 2012), leaving a considerable part of the variance in initial CPS unexplained.

So far, no longitudinal data on the relation between fluid reasoning and CPS development are available. Still, we expected that fluid reasoning would positively predict CPS development because cognitive abilities that develop earlier may facilitate the later development of other cognitive abilities (Kail, 2007). This has been reported for the relation between short-term memory and processing speed with regard to the development of fluid reasoning (Nettelbeck & Burns, 2010). We thus expected that CPS as a complex cognitive skill (Funke, 2010) and its development would be related to fluid reasoning in ways that would be similar to the relations between short-term memory or processing speed and fluid reasoning.

RQ3: Age and Sex Differences in CPS Development

We expected that older students would show better initial CPS performance with regard to age because cognitive abilities increase as students grow older (Stelzl et al., 1995). We also expected that older students would show smaller increases in CPS performance because cognitive development slows down around the ages of

20–22 (Nettelbeck & Burns, 2010). This negatively accelerated developmental pattern (Ritter & Schooler, 2001; Salthouse, 1994), which is strongest at the end of adolescence when fluid abilities such as fluid reasoning or processing speed reach their peak (Nettelbeck & Burns, 2010), would be illustrated by positive relations between age and initial CPS skills and negative relations between age and CPS development.

Regarding sex differences, we expected that initial CPS performance would be better for boys than for girls because previous investigations have reported advantages in CPS in favor of boys (Wittmann & Hattrup, 2004; Wüstenberg et al., 2014). This performance difference has previously been reported for older students' fluid reasoning as well (Lynn & Irwing, 2004), although current research has mostly indicated no sex differences in fluid reasoning (cf. Savage-McGlynn, 2012). We had no specific expectations concerning differences between boys and girls in their CPS development.

Theoretically, there are two possible explanations for sex differences in CPS. On the one hand, boys may apply more efficient strategies when performing CPS tasks. On the other hand, scenario effects may explain males' better performance (Wüstenberg et al., 2014). For instance, Wittmann and Hattrup (2004) reported less risk aversiveness for boys than girls while solving CPS tasks. Thereby, boys generated more opportunities to learn about hidden information and thus showed better performance. However, it may also be possible that the semantic contexts of CPS scenarios, for instance, the cover stories of CPS tasks such as producing and selling shirts in a tailorshop simulation (Danner, Hagemann, Holt, et al., 2011), motivate boys more than girls and thus lead to scenario effects that favor boys (Wüstenberg et al., 2014). Hence, we expected that initial CPS performance would be better for boys than for girls according to theoretical ideas put forth by Wüstenberg et al. (2014) and earlier results from CPS research (Wittmann & Hattrup, 2004).

Method

Participants

The sample used in the present study was taken from three cross-sectional student samples from a Comprehensive School in Southern Germany ($N_{\text{MP1}} = 309$; $N_{\text{MP2}} = 577$, $N_{\text{MP3}} = 551$).[2] The measurement points were approximately 1 year apart, and the students in these samples originated from all three German school tracks (i.e., general, intermediate, and high school education).

From these three cross-sectional samples, only data from students who participated in at least two measurement points were analyzed in the present study. This resulted in a subsample of 277 students (19.1% participated at all three measurement points) with a mean age of 13.62 years ($SD = 1.05$, range = 12–17) at the first measurement point; 56.0% (155) of them were female.[3] In that sample, the composition of students from different school tracks remained the same across all three measurement points (11%–17% from general education, 44%–49% from intermediate education, and 38%–41% from high school education), $\chi^2(4) = 0.526$, $p = .97$. This indicated that the selection of students who participated at a minimum of two measurement points was not biased by an overrepresentation of high-performing students at the second and third measurement points. Still, as the period of data collection lasted for 2 years, and as some older students graduated from high school, the attrition rate was rather high. Students were usually in the eighth or ninth grade at the first measurement occasion and ended the study in the 10th or 11th grade, respectively. Participation was compulsory, and as all students were underage, parents provided informed consent.

Measures

CPS (MicroDYN). CPS was assessed by using tasks based on the MicroDYN approach (Greiff et al., 2012). In the present study, eight separate MicroDYN tasks were used (see Figure 1 for an example). These tasks consisted of two different phases each measuring one of the two main aspects of CPS: knowledge acquisition or knowledge application. First, students were asked to detect causal relations between two to three manipulable input variables and two to three nonmanipulable output variables. Afterwards, students were asked to reach given target values by controlling the variables. Thereby, MicroDYN allows the user to differentiate between the two CPS dimensions that were described theoretically (see CPS section): knowledge acquisition (ACQ) and knowledge application (APP; Greiff, 2012a).

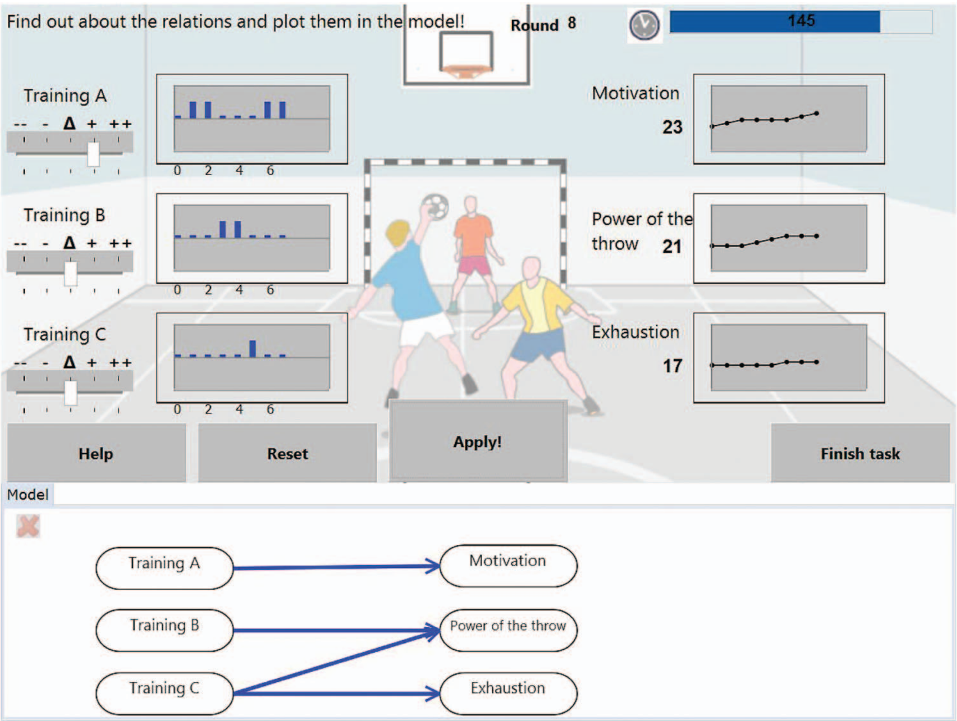
Each of the eight MicroDYN tasks in the present study presented a minimally complex scenario (e.g., coaching a handball team; see Figure 1) with an opaque system structure that was not shown to the students (the structural equations for all tasks are presented in the Appendix). All tasks consisted of two phases. In the first phase, the knowledge acquisition phase, students were instructed to identify the underlying system structure and draw relations between the variables in a model panel given at the bottom of the task (see Figure 1). Thereby, students had to explore the task by altering input variable values (i.e., changing slide controllers; see Figure 1) and detecting corresponding changes in output variable values (see Figure 1).

For example, in the handball training tasks (see Figure 1), students were instructed to figure out how different training methods (on the left side of Figure 1) influenced the training results (on the right side of Figure 1). For that, participants could move the sliders of one or more training methods to a specific value (e.g., "+" for Training A) and could click the *Apply* button to see how their manipulation would influence the training results (e.g., motivation). Simultaneously, students were asked to draw the relations between the inputs and outputs they detected by exploring the task in the model panel given at the bottom of the scenario (see Figure 1). Thus, knowledge acquisition was assessed by an evaluation of the correctness of the model the students drew. The processing time was restricted to 180 s for this phase.

---

[2] Some data from single measurement points were analyzed, and papers reporting these data are currently under revision for publication. These papers compared students from different countries on their CPS skills. However, these papers used only cross-sectional data from either Measurement Points 1 or 2, and none of these analyses included longitudinal analyses across two or all three measurement points.

[3] In order to detect a possible selection bias, CPS skills and fluid reasoning at the first measurement point were compared between students included in the analysis and those who participated only at MP1. CPS and fluid reasoning at the first MP did not differ significantly between these groups of students, $t(308)_{\text{ACQ}} = .347$, $p = .729$; $t(308)_{\text{APP}} = .336$, $p = .737$; $t(308)_{\text{CFT}} = .423$, $p = .672$, indicating no selection bias.



Find out about the relations and plot them in the model!

Figure 1. Screenshot of the MicroDYN Task “handball training” in the knowledge acquisition phase. Controllers of the input variables range from “- -” (value = −2) to “++” (value = +2) and are located on the left side. The current values of the output variables are displayed graphically and numerically on the right side (Wüstenberg et al., 2012). See the online article for the color version of this figure.

In the second phase, the knowledge application phase, students were asked to reach the given target values in the output variables within a maximum of four control steps (e.g., motivation should be between 20 and 22). In this phase, the correct model was given so that, conceptually, control performance did not rely on performance in the preceding knowledge acquisition phase. Knowledge application was assessed by a determination of whether the target values were reached or not. In this phase, the processing time was restricted to 90 s.

To vary the difficulty across the different MicroDYN tasks, the complexity of the underlying system structures of the tasks was gradually increased by adding more input and output variables, by increasing the number of relations between them, and by introducing output variables that changed by themselves over time without active manipulation of the input variables (i.e., autoregressive effects).[4]

Following Greiff, Wüstenberg, and Funke’s (2012) scoring procedure, the scoring of knowledge acquisition and knowledge application was dichotomous. For knowledge acquisition, credit was given if the model the student drew was completely correct; for knowledge application, credit was given if all target values were reached. Thus, for all eight MicroDYN tasks, two dichotomous scores were computed, one for knowledge acquisition and one for knowledge application, respectively. Although the two dimensions ACQ and APP are conceptually different and the MicroDYN tasks are designed in a way that the two scores do not necessarily depend on each other, previous research has reported positive correlations between ACQ and APP (Greiff et al., 2012; Wüstenberg et al., 2012). These results most likely indicate some overlap regarding the cognitive processes underlying ACQ and APP, although it is also possible that this relation was determined by a third variable.

With respect to psychometric quality, MicroDYN provides a reliable (Greiff et al., 2012) and valid assessment of CPS and reflects the theoretical foundation of CPS (Greiff, Fischer, et al., 2013; Molnár et al., 2013; Wüstenberg et al., 2012). Regarding the validity of the MicroDYN approach, different studies have linked MicroDYN to other measurement instruments that are used to assess CPS and to other cognitive abilities such as fluid reasoning or short-term memory (Schweizer et al., 2013; Wüstenberg et al., 2012). The results of these investigations have indicated that MicroDYN converges with other measures of CPS and that MicroDYN differs considerably from measures of other cognitive abilities such as fluid reasoning or short-term memory (Greiff, Fischer, et al., 2013; Greiff, Wüstenberg, et al., 2013; Schweizer et al., 2013; Wüstenberg et al., 2012). Furthermore, MicroDYN has been tested and validated as a measure of CPS in student samples (Greiff & Fischer, 2013; Schweizer et al., 2013; Wüstenberg et al., 2012), and some items were used internationally in the PISA 2012 study.

Fluid reasoning. Fluid reasoning was tested with two nonverbal matrix tests. At the first measurement occasion, all students completed a paper-and-pencil version of the Culture Fair Test

---

[4] For further information and details on varying the difficulty in MicroDYN, see Greiff et al. (2012).

(CFT; Weiß, 2006). At the second measurement occasion, all students completed a computer-based assessment of the nonverbal scale of the Cognitive Abilities Test (KFT; Heller & Perleth, 2000). For the KFT, two items (Items 9 and 14) were not included in further analyses as recent investigations have reported that these two items are unsolvable due to ambiguous solutions (Segerer, Marx, & Marx, 2012).

For the CFT and KFT, a dichotomous score for each item was calculated. Full credit was assigned if the student solved the item correctly; otherwise, no credit was given. Thus, one dichotomous score was computed for each item of each fluid reasoning task.

Procedure

Testing took place in the school's local computer room. Between subsequent measurement points, there was an interval of about 1 year. An entire class composed of 10–25 students was tested in each session.

At each of the three measurement points (MPs), the assessment began with MicroDYN. (The respective MPs will be indicated by index numbers: 1 for the first, 2 for the second, and 3 for the third MP.) Test administration began with general instructions, including one MicroDYN task that the students could use to practice and to become acquainted with the user interface. Subsequently, students worked on eight different MicroDYN tasks. At MP1 and MP2, fluid reasoning was assessed afterwards. Students also provided demographic data at each MP, and some other measures that were not relevant for the present study were assessed as well. Overall, each test session took approximately 90 min.

Statistical Analyses

To analyze the development of CPS, we estimated latent growth curve models (LGCMs) that included both CPS dimensions in one model. LGCMs offer the ability to model differences between MPs through a priori defined growth patterns such as linear or quadratic growth by using structural equation modeling (B. O. Muthén & Khoo, 1998). These models are used to estimate the development of performance and to identify differences in developmental patterns (Duncan, Duncan, & Strycker, 2010).

The latent means of all MPs were estimated in LGCMs by computing regressions with an intercept variable and one or more slope variables (B. O. Muthén & Khoo, 1998). The intercept (I) represents the latent mean of the initial performance (initial CPS), whereas slope variables represent linear (L) or quadratic (Q) growth in relation to this initial performance. Specifically, slope variables represent a pattern of differences between MPs (i.e., development). For example, a mean linear slope value of .10 and a mean intercept of .30 indicate that overall, the mean of the first measurement was .30, and it increased steadily by .10 at each subsequent measurement occasion ($M_1 = .30$, $M_2 = .40$, $M_3 = .50, \ldots$; index numbers indicate the measurement occasion). Thus, a significant mean linear slope indicates linear development, whereas a significant mean quadratic slope indicates quadratic development.

Furthermore, intercept and slope variables were estimated for each individual so that interindividual differences in growth patterns could be assessed (Duncan et al., 2010). This allowed us to calculate correlations between intercept and slope variables and time-invariant predictors (e.g., the prediction of the CPS intercept and the slopes through fluid reasoning, age, and sex).

Identification of the mean structures of the LGCMs was achieved by fixing the intercept of the first parcel to zero for each MP and CPS dimension. As the mean structure is not identified with a completely freed quadratic slope factor in an LGCM with three MPs (Duncan et al., 2010), the variance of the quadratic slope variable was fixed to zero. Thus, only an overall mean for all participants (and no covariance with other measures) was estimated for the quadratic slope.

As the originally dichotomous CPS indicators were parceled (i.e., several indicators were summed to build a parcel; see below) and thus continuous (Little, Cunningham, & Shahar, 2002), full information maximum likelihood (FIML) estimation was used to estimate the means, intercepts, and missing values. Missing values were assumed to be missing at random, thus rendering FIML as the appropriate estimation method for the structural equation modeling analysis (Schafer & Graham, 2002). To evaluate the goodness of fit of the respective models, the chi-square test and several fit indices (comparative fit index [CFI]; Tucker–Lewis Index [TLI]; and root-mean-square error of approximation [RMSEA]) were calculated. As criteria for close model fit, a χ^2 to *df* ratio < 2, CFI and TLI $> .95$, and RMSEA $< .06$ were endorsed (Hu & Bentler, 1999). MPlus 7.0 (L. K. Muthén & Muthén, 1998–2012) was used to compute all models.

Results

Descriptives

The mean CPS performance scores at all three MPs are given in Table 1. Comparisons of these means across MPs indicated a tendency toward better performance at later MPs. Furthermore, the internal consistencies of the two CPS dimensions were good to acceptable, replicating earlier estimations of reliability for MicroDYN (see Table 1). With regard to fluid reasoning, the global score of the CFT at MP1 ($M = 37.3$, $SD = 6.86$, $\alpha = .71$) was comparable to that of the original standardization sample ($M = 36.4$, $SD = 6.56$; Weiß, 2006). For the KFT, performance in our sample ($M = 15.51$, $SD = 5.60$, $\alpha = .89$) was slightly below the performance of the original standardization sample ($M = 17.1$,

Table 1
Descriptive Statistics for Knowledge Acquisition and Knowledge Application and Cronbach's Alpha for the Two CPS Dimensions and Cognitive Tests at All Three Measurement Points

	Descriptive statistics								
	MP1 ($N = 123$)		MP2 ($N = 251$)		MP3 ($N = 226$)		Cronbach's α		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	MP1	MP2	MP3
ACQ	0.38	0.25	0.45	0.26	0.52	0.26	.75	.74	.80
APP	0.27	0.25	0.33	0.28	0.41	0.30	.67	.76	.77
CFT	37.3	6.86					.71		
KFT			15.51	5.60				.89	

Note. CPS = complex problem solving; MP = measurement point; ACQ = knowledge acquisition; APP = knowledge application; CFT = Culture Fair Test; KFT = Cognitive Abilities Test.

SD = 4.66; Heller & Perleth, 2000). The effect was small (d = 0.31), but the range of cognitive ability in the present sample appeared to be representative.

Measurement Models

CPS. For the two dimensions of CPS (i.e., ACQ and APP), the eight tasks were aggregated into three parcels for each MP in order to provide a more parsimonious model for estimating the LGCMs. As the two-dimensional structure of MicroDYN had been validated before (Greiff et al., 2012; Wüstenberg et al., 2012), the application of a parceling procedure was reasonable. For parceling, we used the item-to-construct balance technique (Little et al., 2002), which results in parcels that are balanced with regard to difficulty and parcel discrimination. The measurement models for ACQ and APP were estimated with the loading of the first parcel fixed and the restriction that the two other parcels had to have the same loading due to the measurement-invariance requirements for LGCMs. All measurement models for CPS fit the data very well (see Table 2).

Fluid reasoning. For fluid reasoning, global scores were computed as the means of item scores for the two fluid reasoning measures, resulting in a global score for the CFT and a global score for the KFT. The measurement model for fluid reasoning was estimated with the two global scores of the CFT and KFT as indicators and with both loadings on the latent factor for fluid reasoning fixed to one. Model fit for the measurement model was very good (see Table 2).

RQ1: Development of CPS

To analyze initial CPS and the developmental pattern of CPS, we estimated an LGCM that included both of the dimensions, ACQ and APP. For each dimension, linear (L) and quadratic slopes (Q) were estimated to represent CPS development. Initial CPS performance was represented by the intercepts (I) of both CPS dimensions. The fit indices for the LGCM that included both CPS dimensions were, $\chi^2(149) = 205.715$, $p < .01$, CFI = .962, RMSEA = .037. Although the chi-square test indicated significant differences between the implied and empirical mean and covariance structures, other fit indices such as the CFI and RMSEA indicated a good model fit (Hoelter, 1983; Hu & Bentler, 1999; Thacker, Fields, & Tetrick, 1989). To this end, close inspection of the residual covariance matrix between the implied and actual covariance matrices did not yield any residual correlations greater than .10. Thus, we concluded that the fit of the model was at least acceptable, and we retained the model, which is displayed in Figure 2.

Table 2
Goodness-of-Fit Indices for Measurement Models for Knowledge Acquisition and Knowledge Application at All Three Measurement Points and for Fluid Reasoning Measures

Measurement model	χ^2	*df*	*p*	χ^2/df	CFI	TLI	RMSEA
ACQ (1)	0.410	1	.52	0.41	1.000	1.000	0.000
ACQ (2)	1.023	1	.31	1.02	1.000	1.000	0.010
ACQ (3)	0.029	1	.86	0.03	1.000	1.000	0.000
APP (1)	1.118	1	.29	1.12	0.998	0.995	0.031
APP (2)	0.059	1	.81	0.06	1.000	1.000	0.000
APP (3)	1.638	1	.20	1.64	0.997	0.991	0.053
Fluid reasoning	0.546	1	.46	0.55	1.000	1.000	0.000

Note. χ^2 and *df* are estimated by full information maximum likelihood. CFI = comparative fit index; TLI = Tucker-Lewis Index; RMSEA = root-mean-square error of approximation; ACQ = knowledge acquisition; APP = knowledge application. Parenthetical values indicate the measurement point.

In this model, initial CPS performance at MP1 was moderate for ACQ, indicating considerable potential for development (see Figure 2 for the estimated values). Regarding the development of CPS, the results indicated positive development for both CPS dimensions as expected (effect size of growth for ACQ: from MP1 to MP2: $d = 0.58$, and from MP2 to MP3: $d = 0.41$; for APP: from MP1 to MP2: $d = 0.51$, and from MP2 to MP3: $d = 0.42$). With respect to the shape of the growth, the linear slope was positive for ACQ and APP. By contrast, the quadratic slope was not significantly different from zero for either CPS dimension, indicating no quadratic growth pattern in CPS development (see Figure 2 for the latent means).

The intercepts of ACQ and APP were positively related, indicating that students with good performance in ACQ performed well in APP as well. The relation between the linear slope variables of ACQ and APP was also positive, so students with large increases in ACQ had large increases in APP, too (see Figure 2 for the correlations).

The correlations between the intercept and linear slope variables were significant and negative for both CPS dimensions (see Figure 2), thus indicating that students showing better performance at the first MP had smaller increases in CPS performance at later MPs. Because we had to fix the variance of the quadratic slope to zero (see Statistical Analyses section), no covariance of the quadratic slope with any other variable was estimated.

RQ2: Explaining CPS Development Through Fluid Reasoning

In a second step (to test RQs 2 and 3), we additionally included fluid reasoning, sex, and age as predictors of the intercepts and slopes in our estimations of the LGCMs for the two CPS dimensions. To report the results in a straightforward manner, the relations between fluid reasoning and the CPS variables will be presented first to address RQ2, and the relations between age and sex and the CPS variables will be presented afterwards to address RQ3. For the complete LGCM with all predictors standardized, see Figure 3.

The overall fit of the LGCM with all covariates was good, $\chi^2(205) = 261.111$, $p < .01$, CFI = .965, RMSEA = .031. Again, the chi-square test indicated significant differences between the implied and empirical mean and covariance structures, but the other fit indices suggested a good fit (see Statistical Analyses section). Thus, we retained the model. The parameters of the LGCMs (intercepts, linear, and quadratic slopes) for both CPS dimensions showed little change in the models with fluid reasoning, age, and sex as predictors compared with the LGCMs without the predictors presented in Figure 2. Only the relation between the intercept and the respective linear slope variable of both CPS dimensions increased substantially in the model with fluid reasoning, age, and sex (see Figure 3).

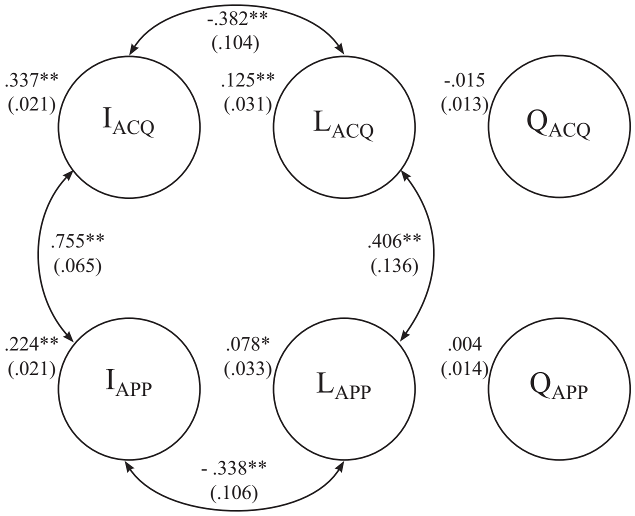


Figure 2. Latent growth curve models displaying the relations between the two complex problem solving dimensions, ACQ and APP. Numbers at the top left of the latent variables represent the latent means; numbers in parentheses indicate standard errors. Correlations between the intercept of ACQ and the linear slope of APP, and vice versa, are not displayed as they failed to reach statistical significance. Correlations with the quadratic slope variables and any other variable were not estimated because the variance of the quadratic slope was fixed to zero. I = intercept; L = linear slope variable; Q = quadratic slope variable; ACQ = knowledge acquisition; APP = knowledge application. * $p < .05$. ** $p < .01$.

With respect to the prediction of initial CPS (left side of Figure 3), fluid reasoning significantly positively predicted the intercepts of both CPS dimensions, indicating that students with better fluid reasoning showed better initial performance on both CPS dimensions. Regarding the prediction of CPS development, the linear slopes of both CPS dimensions were significantly positively related to fluid reasoning, indicating greater CPS development for students with better fluid reasoning ability. However, the regression of the linear slope of ACQ on fluid reasoning failed to reach statistical significance due to a large standard error of the estimate (left side of Figure 3). Nevertheless, the size of the relation indicates that there might be a positive relation between ACQ and fluid reasoning. In summary, fluid reasoning predicted initial CPS and CPS development. However, initial ACQ performance was not significantly predicted by fluid reasoning due to a large standard error.

To address the crucial question of whether CPS relies on fluid reasoning and not vice versa, we conducted two additional latent regression analyses: In the first analysis, the two CPS dimensions (i.e., ACQ and APP) measured at MP2 and MP3 were predicted by fluid reasoning scores measured at MP1 (CFT). In the second analysis, fluid reasoning scores at MP2 (KFT) were predicted by ACQ and APP at MP1. The fit of the regression model in Analysis 1 was acceptable, $\chi^2(62) = 93.123$, $p < .01$, RMSEA = .064, CFI = .934. In this analysis, CFT scores at MP1 significantly predicted ACQ and APP at MP2 and at MP3 (see left side of Figure 4). The second regression model fit the data well, $\chi^2(14) = 15.936$, $p = .32$, RMSEA = .022, CFI = .991. In this second analysis, only ACQ at MP1 predicted KFT scores at MP2, whereas APP at MP1 yielded no substantial prediction of KFT scores at MP2 (see right side of Figure 4). Thus, these additional analyses indicated that CPS at later MPs relied on fluid reasoning, whereas fluid reasoning at later MPs was predicted only by ACQ and not by APP skills from an earlier MP. Also, fluid reasoning at MP2 (i.e., KFT) relied on ACQ at MP1 to a lesser extent than fluid reasoning at MP1 predicted ACQ at MP2 ($z = 2.53$, $p < .05$).

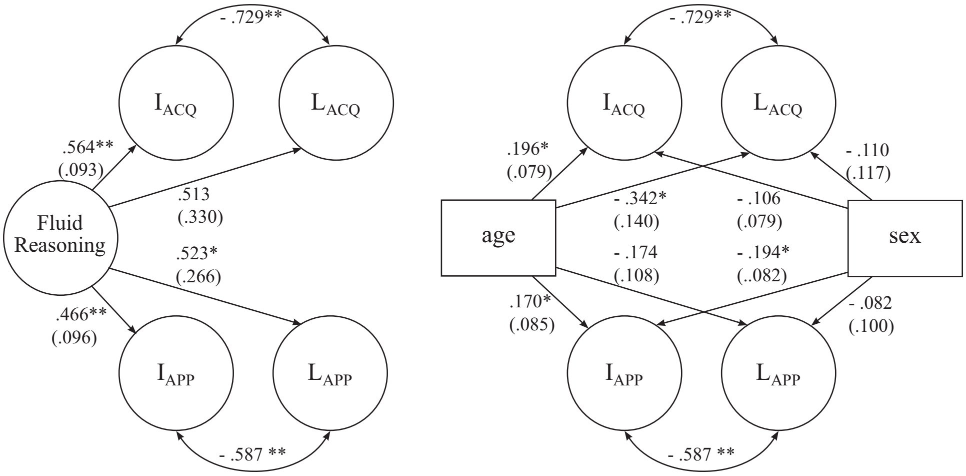


Figure 3. Latent growth curve models of both complex problem solving (CPS) dimensions with fluid reasoning, age, and sex as predictors. Left side: Fluid reasoning predicted initial CPS (I) and the linear growth of CPS (L); right side: Age and sex predicted I and L; all path coefficients were estimated within one model. Quadratic slopes (Q) are not displayed as there was no significant quadratic development in CPS for either dimension. Paths from sex and age to fluid reasoning were not displayed as both paths were not significant. I = intercept; L = linear slope; ACQ = knowledge acquisition; APP = knowledge application. Regression coefficients are standardized. Numbers in parentheses indicate standard errors. * $p < .05$. ** $p < .01$.

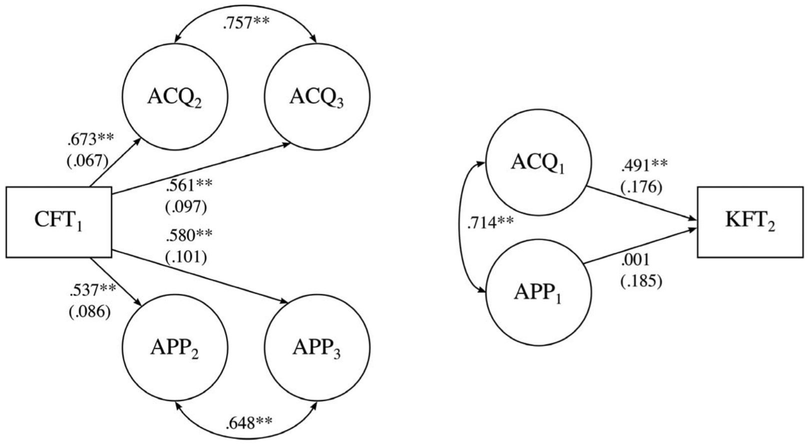


Figure 4. Left part (Analysis 1): CFT scores at Measurement Point (MP) 1 predict CPS scores for knowledge acquisition (ACQ) and knowledge application (APP) at MP2 and MP3 (indicated by the index numbers). Right part (Analysis 2): CPS scores at MP1 predict KFT scores measured at MP2. All coefficients are standardized. Values in parentheses indicate standard errors. CFT = Culture Fair Test; KFT = Cognitive Abilities Test. ** $p <$ .01.

RQ3: Age and Sex Differences in CPS Development

With regard to the prediction of initial CPS by age (right side of Figure 3), the regression of the intercept of CPS on age indicated a significant positive effect of age on both CPS dimensions such that older students performed better. Furthermore, age predicted the linear slopes of both CPS dimensions, indicating smaller increases in CPS performance for older students, a finding that is in line with the expected negatively accelerated developmental pattern. However, the regression of the linear slope of APP on age failed to reach statistical significance (see standard errors in Figure 3, right side). In summary, we found that the CPS intercept was positively related to age, whereas the linear slope of ACQ was negatively related to age, and the linear slope of APP was not related to age.

With regard to sex differences in initial CPS and CPS development (right side of Figure 3), sex did not significantly predict the intercept of ACQ, whereas sex was a significant predictor of the intercept of APP, indicating better performance for boys. Furthermore, sex did not significantly predict the slope variables of either CPS dimension. Thus, boys showed better initial APP performance, whereas no sex differences were found for initial ACQ performance and the slopes of CPS.

Discussion

The present study focused on the longitudinal development of CPS in students ranging in age from 12 to 17 with the aims of determining the developmental patterns of CPS (RQ1), investigating whether fluid reasoning would predict initial CPS and the development of CPS (RQ2), and searching for age and sex differences in CPS and the development of CPS (RQ3). The results indicated the positive linear growth of CPS across MPs. Further, fluid reasoning predicted initial CPS as well as the development of CPS. Age was positively related to initial CPS and negatively related to the development of knowledge acquisition, whereas age did not predict the development of knowledge application. Finally, sex did not predict initial CPS, except for knowledge application, and had no effect on the development of CPS (RQ3).

Development of CPS

With respect to the development of CPS, the results suggested a positive linear growth of CPS with medium effect sizes, not only cross-sectionally, as found by Molnár et al. (2013), but longitudinally as well. Thus, the present results further validate developmental patterns in students' CPS obtained in cross-sectional samples (Greiff, Wüstenberg, et al., 2013; Molnár et al., 2013). Moreover, Stelzl et al. (1995) reported a linear development of cognitive abilities such as fluid reasoning. This pattern of development was illustrated for CPS in our results as well, indicating that the developmental pattern of CPS is similar to that of other cognitive abilities such as fluid reasoning.

Regarding the relation between the two CPS dimensions, a positive correlation was found between initial performance in knowledge acquisition and initial performance in knowledge application. Development in one dimension was also correlated with development in the other. The positive correlation between initial CPS skills replicated findings from earlier investigations (Schweizer et al., 2013; Wüstenberg et al., 2012) pointing toward CPS as a general skill with two correlated but separable dimensions. Furthermore, the medium correlation between the development of the two CPS dimensions indicated that knowledge acquisition and knowledge application develop partly independently rather than in complete correspondence.

The development of each CPS dimension was negatively related to the corresponding measure of initial performance. Thus, smaller increases at subsequent MPs were found for students who performed well at the first MP. This might be due to better learning opportunities on easier CPS tasks, resulting in a negatively accelerated learning curve (Ritter & Schooler, 2001; Salthouse, 1994) and indicating that advances on easier CPS tasks were accomplished more quickly than on more difficult CPS tasks that contained, for instance, nontransparent developments without active user interventions (i.e., indirect effects; Greiff et al., 2012). However, there might be other explanations for the observed pattern such as ceiling effects. Nevertheless, the negative relations that we found between initial performance in CPS and CPS development have important implications for further research, as they indicate that students with moderate or low CPS skills may reap the largest benefits from interventions aimed at fostering their CPS skills.

Fluid Reasoning as a Predictor of CPS Development

Initial CPS performance was positively related to fluid reasoning, thus replicating earlier findings. However, as also reported in previous studies (Greiff, Wüstenberg, et al., 2013; Molnár et al., 2013; Wüstenberg et al., 2012), considerable parts of the variance of CPS could not be explained by fluid reasoning. Thus, CPS and fluid reasoning share some important demands (Wiley et al., 2011) but differ in other core features such as interactivity and dynamics, both of which are central aspects of CPS tasks (Raven, 2000; Wüstenberg et al., 2012).

With respect to predicting CPS development with fluid reasoning, we found in the present investigation that fluid reasoning predicted the development of both CPS dimensions. Although the prediction of the development of knowledge acquisition failed to reach statistical significance ($p = .060$), the size of the regression weight ($\beta = .513$) indicated that at least some variance in the development of knowledge acquisition might be explained by fluid reasoning. To this end, previous research has reported that early developing cognitive abilities such as early executive functioning, short-term memory, or processing speed predict the development of later developing cognitive abilities such as fluid reasoning (Fry & Hale, 1996; Nettelbeck & Burns, 2010; Richland & Burchinal, 2013). Regarding the theoretical conceptualization of CPS development and empirical relations between CPS development and fluid reasoning, a corresponding association between CPS and fluid reasoning was expected. As fluid reasoning at the first measurement point predicted CPS skills at the following MPs (βs = .537–.673), whereas only knowledge acquisition at the first MP predicted fluid reasoning at the second MP ($\beta = .491$, see Results on RQ2), the expected developmental cascade from fluid reasoning to CPS was validated. This further supports the theoretical perspective on cognitive development as a cascade that advances from early developing cognitive abilities up to the later development of complex cognitive skills such as CPS (Fry & Hale, 1996; Nettelbeck & Burns, 2010).

Age and Sex Differences in CPS Development

Regarding age differences in initial CPS and CPS development, in line with recent studies (Greiff, Wüstenberg, et al., 2013; Molnár et al., 2013), our results indicate that initial CPS is positively related to age. The results of the present study further indicate that CPS development slows down as students grow older, thus describing a negatively accelerated developmental pattern toward the end of cognitive development. This has also been reported for other cognitive abilities (Nettelbeck & Burns, 2010). With regard to the time span in which CPS develops, the present results indicate that CPS skills are mostly fostered in the seventh and eighth grades (Molnár et al., 2013). This finding points toward the importance of facilitating CPS development during this period through formal and informal learning approaches that emphasize the relevance of adequate training opportunities for students at this particular age.

For the relation between sex and CPS, the present results suggest that CPS performance differs between boys and girls only with regard to initial performance in knowledge application. Beyond this, there were no sex differences in the development of the two CPS dimensions. Still, there was an advantage for boys compared with girls in their initial level of knowledge application. This result supports the theoretical idea that boys apply more suitable strategies when solving CPS tasks (Wüstenberg et al., 2014), and this may ultimately result in a higher performance in knowledge application. However, the effect size was much smaller in this study compared with earlier results (Wittmann & Hattrup, 2004; Wüstenberg et al., 2014). Even further, there were no sex differences in CPS development or in the initial level of knowledge acquisition in the remaining results. These results were not as expected and contradict the results of earlier investigations on sex differences in CPS (Wittmann & Hattrup, 2004; Wüstenberg et al., 2014). Regarding educational implications, our results, especially the results that indicated no sex differences in CPS development, may be important as a great deal of effort has been put forth in recent decades to provide boys and girls with adequate educational opportunities so that both can develop their cognitive abilities as much as possible (Bailey, 1993).

Limitations of the Present Study

The present study has some limitations that we would like to point out. First, it is not possible to distinguish genuine growth or development in CPS skills from practice effects on the specific CPS tasks within the present framework and design of the study. This is a problem in all longitudinal investigations assessing the same measures at different MPs (Roediger & Karpicke, 2006). However, the interval between the MPs was rather long in this study (i.e., 9–12 months) compared with other studies (Jang, Wixted, Pecher, Zeelenberg, & Huber, 2011; Theisen, Rapport, Axelrod, & Brines, 1998), and practice effects decrease with longer intervals between MPs (Jang et al., 2011; Theisen et al., 1998). Thus, it is unlikely that increases in CPS skills can be completely accounted for by practice effects. Still, for more reliable insights into the development of CPS, replications of the present results with experimental and control group designs are needed (B. O. Muthén & Curran, 1997).

Second, we had to fix the variance of the quadratic slope to zero (see the Statistical Analyses section) so that only one general mean of the quadratic slope was estimated for all participants. Thus, we were able to investigate only whether there was quadratic development in general. Although the quadratic slope was zero overall, some participants may have had a negative and some a positive quadratic slope, which would thus result in an overall quadratic slope of zero. This could not be tested in the present study. Furthermore, we were not able to test how these differences in quadratic development might be related to other model parameters such as initial CPS skills, fluid reasoning, sex, or age.

Moreover, we observed substantial attrition in our sample, resulting in a rather small number of students in relation to the complex data analyses that we conducted. This is a common problem in longitudinal investigations (Ibrahim & Molenberghs, 2009). However, we took several steps to ensure that this attrition was not selective with regard to the cognitive abilities of the students we included.

Additionally, age effects on CPS as well as other measures (e.g., fluid reasoning) were rather small or nonsignificant in the present study. Such findings are rather uncommon (e.g., Savage-McGlynn, 2012). However, the age range was rather narrow, with usually only 2 years between the youngest and the oldest participant and only a few participants who were 3 or more years older than the youngest participants at all MPs. Thus, it is likely that, in general, age effects were underestimated in the present study and that the small and nonsignificant effects of age on CPS and other measures such as fluid reasoning were due to a restriction in age.

Finally, some parameter estimates had surprisingly large standard errors, resulting in large regression weights that were not statistically significant (e.g., the regression of ACQ on fluid reasoning). This indicates a lack of power in the present analyses. Thus, further research is necessary to more closely evaluate the effect sizes and the significance of the relations between CPS, fluid reasoning, sex, and age. Besides, to fully estimate a quadratic slope and its variance, further longitudinal studies with at least four measurement occasions are needed.

Conclusion

The present investigation provides the first insights into the development of CPS and the relations between CPS development and fluid reasoning, age, and sex. Previous studies on cognitive development have reported that cognitive abilities that develop early often enhance the later development of other cognitive abilities (Kail, 2007; Nettelbeck & Burns, 2010). As CPS development seems to rely at least partly on fluid reasoning, CPS may reasonably be integrated into this cascade of cognitive development (Fry & Hale, 1996). However, fluid reasoning still does not fully explain CPS (Greiff, Fischer, et al., 2013; Wüstenberg et al., 2012) and its development. Although it is possible that CPS development is enhanced through training methods that aim to enhance fluid reasoning (Molnár et al., 2013), the specific means for enhancing CPS skills are scarce at best. The results of the present study may provide a foundation for further research investigating methods that enhance the development of CPS skills.

Mayer and Wittrock (2006) stated that CPS in general, and, more specifically, the enhancement of CPS development, is one of the most important topics in education, emphasizing that analytical, interactive, and nonroutine tasks are becoming increasingly important in today's world (Autor et al., 2003). Even further, the OECD has summarized the importance of such skills and their implementation in educational systems all over the world (OECD, 2010). To conclude, the facilitation of CPS is crucial because today's world is changing quickly, and the ability to approach new and unfamiliar situations is of growing importance. Thus, we need to understand how complex cognitive abilities progress in education and how their development can be enhanced to specifically influence these abilities within the educational process. This is essential for providing children and students with educations that are appropriate for the 21st century.

References

Autor, D. H., Levy, F., & Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics, 118,* 1279–1333. doi:10.1162/003355303322552801

Babcock, R. L. (2002). Analysis of age differences in types of errors on the Raven's Advanced Progressive Matrices. *Intelligence, 30,* 485–503. doi:10.1016/S0160-2896(02)00124-1

Bailey, S. M. (1993). The current status of gender equity research in American schools. *Educational Psychologist, 28,* 321–339. doi:10.1207/s15326985ep2804_3

Berninger, V. W., & Abbott, R. D. (2010). Listening comprehension, oral expression, reading comprehension, and written expression: Related yet unique language systems in Grades 1, 3, 5, and 7. *Journal of Educational Psychology, 102,* 635–651. doi:10.1037/a0019319

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. 27–63). Hillsdale, NJ: Erlbaum.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves from an early childhood educational experiment. *Developmental Psychology, 37,* 231–242.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97,* 404–431. doi:10.1037/0033-295X.97.3.404

Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 69–76). New York, NY: Guilford Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin.

Coyle, T. R., Pillow, D. R., Snyder, A. C., & Kochunov, P. (2011). Processing speed mediates the development of general intelligence (g) in adolescence. *Psychological Science, 22,* 1265–1269. doi:10.1177/0956797611418243

Crosnoe, R., Morrison, F., Burchinal, M., Pianta, R., Keating, D., Friedman, S. L., & Clarke-Stewart, K. A. (2010). Instruction, teacher–student relations, and math achievement trajectories in elementary school. *Journal of Educational Psychology, 102,* 407–417. doi:10.1037/a0017762

Danner, D., Hagemann, D., Holt, D. V., Hager, M., Schankin, A., Wüstenberg, S., & Funke, J. (2011). Measuring performance in dynamic decision making: Reliability and validity of the tailorshop simulation. *Journal of Individual Differences, 32,* 225–233. doi:10.1027/1614-0001/a000055

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond iq: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence, 39,* 323–334. doi:10.1016/j.intell.2011.06.004

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35,* 13–21. doi:10.1016/j.intell.2006.02.001

de Jong, P., & de Jong, P. F. (1996). Working memory, intelligence and reading ability in children. *Personality and Individual Differences, 21,* 1007–1020. doi:10.1016/S0191-8869(96)00161-4

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatonmy, and bio-chemistry. In D. T. Stuss & R. T. Knught (Eds.), *Principles of frontal lobe function* (pp. 466–503). London, England: Oxford University Press. doi:10.1093/acprof:oso/9780195134971.003.0029

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2010). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Erlbaum.

Fischer, A., Greiff, S., & Funke, J. (2011). The process of solving complex problems. *Journal of Problem Solving, 4,* 19–42. doi:10.7771/1932-6246.1118

Fry, A. F., & Hale, S. (1996). Processing speed, working memory and fluid intelligence. *Psychological Science, 7,* 237–241. doi:10.1111/j.1467-9280.1996.tb00366.x

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning, 7,* 69–89. doi:10.1080/13546780042000046

Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing, 11,* 133–142. doi:10.1007/s10339-009-0345-0

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). The relationships between cognitive ability and dynamic decision making. *Intelligence, 33,* 169–186. doi:10.1016/j.intell.2004.10.002

Greiff, S. (2012a). Assesment and theory in complex problem solving: A continuing contradiction? *Journal of Educational and Developemtnal Psychology, 2,* 49–56.

Greiff, S. (2012b). From interactive to collaborative problem solving: Current issues in the Programme for International Student Assessment. *Review of Psychology, 19,* 111–121.

Greiff, S., & Fischer, A. (2013). Der nutzen eine komplexen problemlösekompetenz: Theoretische überlegungen und empirische befunde [The usefulness of complex problem solving skills: Theoretical considerations and empirical findings]. *Zeitschrift für Pädagogische Psychologie, 27,* 27–39. doi:10.1024/1010-0652/a000086

Greiff, S., Fischer, A., Wüstenberg, S., Sonnleitner, P., Brunner, M., & Martin, R. (2013). A multitrait–multimethod study of assessment instruments for complex problem solving. *Intelligence, 41,* 579–596. doi:10.1016/j.intell.2013.07.012

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic problem solving: A new measurement perspective. *Applied Psychological Measurement, 36,* 189–213. doi:10.1177/0146621612439620

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond *g*: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology, 105,* 364–379. doi:10.1037/a0031856

Heller, K. A., & Perleth, C. (Eds.). (2000). *Kognitiver Fähigkeitstest für 4. Bis 12. Klassen, revision* [Cognitive Abilities Test for 4th to 12th grade, revision]. Göttingen, Germany: Beltz-Test GmbH.

Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research, 11,* 325–344. doi:10.1177/0049124183011003003

Hooper, F. H., Fitzgerald, J., & Papalia, D. (1971). Piagetian theory and the aging process: Extensions and speculations. *Aging & Human Development, 2,* 3–20. doi:10.2190/AG.2.1.b

Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 41–68). New York, NY: Guilford Press.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53–91). New York, NY: Guilford Press.

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

Ibrahim, J. G., & Molenberghs, G. (2009). Missing data methods in longitudinal studies: A review. *TEST, 18,* 1–43. doi:10.1007/s11749-009-0138-x

Jang, Y., Wixted, J. T., Pecher, D., Zeelenberg, R., & Huber, D. E. (2012). Decomposing the interaction between retention interval and study/test practice: The role of retrievability. *The Quarterly Journal of Experimental Psychology, 65,* 962–975. doi:10.1080/17470218.2011.638079

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences*. Oxford, England: Elsevier Science.

Kail, R. V. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109,* 490–501. doi:10.1037/0033-2909.109.3.490

Kail, R. V. (1992). Processing speed, speech rate, and memory. *Developmental Psychology, 28,* 899–904. doi:10.1037/0012-1649.28.5.899

Kail, R. V. (2007). Longitudinal evidence that increases in processing speed and working memory enhance children's reasoning. *Psychological Science, 18,* 312–313. doi:10.1111/j.1467-9280.2007.01895.x

Klahr, D. (Ed.). (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge, MA: MIT Press.

Kröner, S., Plass, J. L., & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence, 33,* 347–368. doi:10.1016/j.intell.2005.03.002

Little, T. D., Cunningham, W. A., & Shahar, G. (2002). To parcel or not to parcel: Exploring the questions, weighing the merits. *Structural Equation Modeling, 9,* 151–173. doi:10.1207/S15328007SEM0902_1

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32,* 481–498. doi:10.1016/j.intell.2004.06.008

Mayer, R. E., & Wittrock, M. C. (2006). Problem solving. In P. A. Alexander & P. H. Winnie (Eds.), *Handbook of educational psychology* (pp. 287–303). Mahwah, NJ: Erlbaum.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1–10. doi:10.1016/j.intell.2008.08.004

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity, 9,* 35–45. doi:10.1016/j.tsc.2013.03.002

Muthén, B. O., & Curran, P. J. (1997). General longitudinal modeling of individual differences in experimental designs: A latent variable framework for analysis and power estimation. *Psychological Methods, 2,* 371–402. doi:10.1037/1082-989X.2.4.371

Muthén, B. O., & Khoo, S.-T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences, 10,* 73–101. doi:10.1016/S1041-6080(99)80135-6

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (Vol. 7). Los Angeles, CA: Author.

Nettelbeck, T., & Burns, N. R. (2010). Processing speed, working memory and reasoning ability from childhood to old age. *Personality and Individual Differences, 48,* 379–384. doi:10.1016/j.paid.2009.10.032

Organization for Economic Co-Operation and Development. (2004). *Problem solving for tomorrow's world. First measures of cross-curricular competencies from Pisa 2003*. Paris, France: Author.

Organization for Economic Co-Operation and Development. (2010). *PISA 2012 problem solving framework (draft for field trial)*. Paris, France: Author.

Raven, J. (2000). Psychometrics, cognitive ability and occupational performance. *Review of Psychology, 7,* 51–74.

Richland, L. E., & Burchinal, M. R. (2013). Early executive function predicts reasoning development. *Psychological Science, 24,* 87–92. doi:10.1177/0956797612450883

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In N. J. Smelser & P. B. Baltes (Eds.), *International encyclopedia of the social & behavioral sciences* (pp. 8602–8605). Oxford, England: Pergamon.

Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181–210. doi:10.1111/j.1745-6916.2006.00012.x

Romine, C. B., & Reynolds, C. R. (2005). A model of the development of frontal lobe functioning: Findings from a meta-analysis. *Applied Neuropsychology, 12,* 190–201. doi:10.1207/s15324826an1204_2

Sackett, P. R., Bornemann, M. J., & Conelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63,* 215–227. doi:10.1037/0003-066X.63.4.215

Salthouse, T. A. (1994). Aging associations: Influence of speed on adult age differences in associative learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 1486–1503. doi:10.1037/0278-7393.20.6.1486

Savage-McGlynn, E. (2012). Sex differences in intelligence in younger and older participants of the Raven's Standard Progressive Matrices plus. *Personality and Individual Differences, 53,* 137–141. doi:10.1016/j.paid.2011.06.013

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. doi:10.1037/1082-989X.7.2.147

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274. doi:10.1037/0033-2909.124.2.262

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences, 24,* 42–52. doi:10.1016/j.lindif.2012.12.011

Segerer, R., Marx, A., & Marx, P. (2012). Unlösbare items im KFT 4–12+r [Unsolvable items in the KFT 4-12+R]. *Diagnostica, 58,* 45–50. doi:10.1026/0012-1924/a000057

Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence, 41,* 289–305. doi:10.1016/j.intell.2013.05.002

Stelzl, I., Merz, F., Ehlers, T., & Remer, H. (1995). The effect of schooling on the development of fluid and cristallized intelligence: A quasi-experimental study. *Intelligence, 21,* 279–296. doi:10.1016/0160-2896(95)90018-7

Stuss, D. T. (2007). New approaches to prefrontal lobe testing. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (2nd ed., pp. 292–305). New York, NY: Guilford Press.

Thacker, J. W., Fields, M. W., & Tetrick, L. E. (1989). The factor structure of union commitment: An application of confirmatory factor analysis. *Journal of Applied Psychology, 74,* 228–232. doi:10.1037/0021-9010.74.2.228

Theisen, M. E., Rapport, L. J., Axelrod, B. N., & Brines, D. B. (1998). Effects of practice in repeated administrations of the Wechsler Memory Scale-Revised in normal adults. *Assessment, 5,* 85–92. doi:10.1177/107319119800500110

Thibaut, J.-P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology, 106,* 1–19. doi:10.1016/j.jecp.2010.01.001

Vaessen, A., Bertrand, D., Tóth, D., Csépe, V., Faísca, L., Reis, A., & Blomert, L. (2010). Cognitive development of fluent word reading does not qualitatively differ between transparent and opaque orthographies. *Journal of Educational Psychology, 102,* 827–842. doi:10.1037/a0019465

Weiß, R. H. (2006). *Grundintelligenztest Skala 2 - Revision (CFT 20-r)* [Culture Fair Test Scale 2-Revised (CFT 20-R)]. Göttingen, Germany: Hogrefe.

Wenke, D., Frensch, P. A., & Funke, J. (2005). Complex problem solving and intelligence: Empirical relation and causal direction. In R. J. Sternberg & J. E. Pretz (Eds.), *Cognition and intelligence: Identifying the mechanisms of the mind* (pp. 160–187): Cambridge, England: Cambridge University Press.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colfiesh, G. J. H. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 256–263. doi:10.1037/a0021613

Wittmann, W. W., & Hattrup, K. (2004). The relationship between performance in dynamic systems and intelligence. *Systems Research and Behavioral Science, 21,* 393–409. doi:10.1002/sres.653

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence, 40,* 1–14. doi:10.1016/j.intell.2011.11.003

Wüstenberg, S., Greiff, S., Molnár, G., & Funke, J. (2014). Cross-national gender differences in complex problem solving and their determinants. *Learning and Individual Differences, 29,* 18–29. doi:10.1016/j.lindif.2013.10.006

(*Appendix follows*)

Appendix

The Eight MicroDYN Tasks Used in the Present Study

Variable	Linear structure equations	System size	Effects
Item 1	$X_{t\,+\,1} = 1^{*} X_t + 0^{*} A_t + 2^{*} B_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 0^{*} A_t + 2^{*} B_t$	2 × 2 system	Only direct
Item 2	$X_{t\,+\,1} = 1^{*} X_t + 2^{*} A_t + 2^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 0^{*} A_t + 0^{*} B_t + 2^{*} C_t$	2 × 3 system	Only direct
Item 3	$X_{t\,+\,1} = 1^{*} X_t + 0^{*} A_t + 2^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Z_{t\,+\,1} = 1^{*} Z_t + 0^{*} A_t + 0^{*} B_t + 2^{*} C_t$	3 × 3 system	Only direct
Item 4	$X_{t\,+\,1} = 1^{*} X_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 0^{*} A_t + 2^{*} B_t + 2^{*} C_t$ $Z_{t\,+\,1} = 1^{*} Z_t + 0^{*} A_t{}^{*}\, 0^{*} B_t + 2^{*} C_t$	3 × 3 system	Only direct
Item 5	$X_{t\,+\,1} = 1^{*} X_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 0^{*} A_t + 2^{*} B_t + 0^{*} C_t$ $Z_{t\,+\,1} = 1^{*} Z_t + 0^{*} A_t{}^{*}\, 2^{*} B_t + 2^{*} C_t$	3 × 3 system	Only direct
Item 6	$X_{t\,+\,1} = 1^{*} X_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1.33^{*} Y_t + 0^{*} A_t + 2^{*} B_t + 0^{*} C_t$	2 × 3 system	Direct and Indirect
Item 7	$X_{t\,+\,1} = 1^{*} X_t + 0^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1.33^{*} Y_t + 2^{*} A_t + 2^{*} B_t + 0^{*} C_t$ $Z_{t\,+\,1} = 1^{*} Z_t + 0^{*} A_t + 0^{*} B_t + 2^{*} C_t$	3 × 3 system	Direct and Indirect
Item 8	$X_{t\,+\,1} = 1^{*} X_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Y_{t\,+\,1} = 1^{*} Y_t + 2^{*} A_t + 0^{*} B_t + 0^{*} C_t$ $Z_{t\,+\,1} = 1.33^{*} Z_t + 0^{*} A_t{}^{*}\, 0^{*} B_t + 2^{*} C_t$	3 × 3 system	Direct and Indirect

Note. The eight MicroDYN tasks used in this study varied in their underlying system structure. For more information on the rationale behind the MicroDYN tasks, see Greiff et al. (2012) and Wüstenberg et al. (2012). X_t, Y_t, and Z_t represent values of the output variables; A_t, B_t, and C_t represent values of the input variables during the present execution round, whereas X_{t+1}, Y_{t+1}, and Z_{t+1} represent values of the output variables in the subsequent execution round.

Received June 7, 2013
Revision received April 26, 2014
Accepted April 27, 2014 ■