9-15-2014

Metacognitive planning: Development and validation of an online measure.

Jian Li
Beijing Normal University

Bo Zhang
Beijing Normal University

Han Du
University of Notre Dame

Ze Zhu
University of Nebraska at Omaha, zezhu@unomaha.edu

Yi Ming Li
Beijing Normal University

Recommended Citation

Li, J., Zhang, B., Du, H., Zhu, Z., & Li, Y. M. (2015). Metacognitive planning: Development and validation of an online measure. Psychological Assessment, 27(1), 260–271. https://doi.org/10.1037/pas0000019

Metacognitive Planning: Development and Validation of an Online Measure

Jian Li
Beijing Normal University and Beijing Key Laboratory of Applied Experimental Psychology, Beijing, China
Bo Zhang
Beijing Normal University
Han Du
University of Notre Dame
Ze Zhu and Yi Ming Li
Beijing Normal University

Planning is the critical first stage of metacognition. Although it has long been emphasized theoretically, it has not been the subject of much empirical study due to the lack of a valid assessment tool. Because planning is a metacognitive process, online methods that collect data during task performance would much better capture it. The present study was conducted to develop an online measure of metacognitive planning. Researchers designed a puzzle task that took the form of the popular game Sokoban, and the ratio between planning time and total time of each item was chosen as the metacognitive planning index. The task was administered to a heterogeneous sample of 440 participants composed of college students as well as 5th-, 7th-, and 10th-grade students. The results showed that valid inference could be made from the time ratio score. Cronbach's alpha and test–retest correlation provided robust evidence of reliability of the time ratio score. Confirmatory factor analysis further confirmed its unidimensionality. Validity evidence also supported the use of the time ratio score. After controlling for demographic variables, intelligence, and motivation, the time ratio score still accounted for a significant proportion of variance of Sokoban performance, the Tower of London performance, and academic achievement. The time ratio score was also found to increase with age. Taken together, the results of the study revealed that the time ratio is a psychometrically sound online measure of metacognitive planning.

***Keywords**:* metacognitive planning, online, puzzle task, time ratio

This research was supported by Ministry of Education in China Project of Humanities and Social Sciences Project No. 09YJCXLX001. The authors declare that they have no competing interests. We owe our sincere thanks to Nian Zu Chen from Beijing No. 8 High School, Xiang Qin Shen from The Affiliated High School of Beijing Normal University, and Jing Yi Liu from Chaoyang Branch of Beijing Institute of Education, without whose help the present research would not be carried out so fluently. Also, we deeply appreciate the help of Xinxian Hutong Elementary School, Beijing Chaoyang Wangjing Experimental School, and Tuan Jie Hu High School.

Planning plays an important role in the field of intelligence (Das, Naglieri, & Kirby, 1994; Naglieri, 2011), executive function (Garden, Phillips, & MacPherson, 2001; Willcutt, Doyle, Nigg, Faraone, & Pennington, 2005; Zelazo, Carter, & Reznick, 1997), and metacognition (Desoete, 2008; Flavell, 1979; Zimmerman, 1989). In the field of intelligence, Luria (1970) believed that three main functional systems or blocks represented the brain's basic functions: (a) Block 1, which is responsible for arousal and attention; (b) Block 2, which is responsible for the analysis, coding, and storage of information; (c) Block 3, which is responsible for the formulation and implementation of plans. Inspired by Luria's work, Naglieri and Das (1997) proposed the Planning, Attention, Simultaneous processing, Successive processing (PASS) theory of intelligence and designed the Cognitive Assessment System (CAS; Naglieri, 2011; Naglieri & Das, 1988, 1990) to assess intelligence. In the same vein, the Kaufmans constructed the second edition of the Kaufman Assessment Battery for Children (KABC-II; A. S. Kaufman & Kaufman, 2004a; J. C. Kaufman, Kaufman, Kaufman- Singer, & Kaufman, 2005). In both tests, *planning* is prioritized as one of the core components of intelligence and defined as "a mental process by which the individual determines, selects, applies and evaluates solutions to problems" (Naglieri & Das, 1997, p. 17).

When approached from the perspective of executive function, planning is commonly conceptualized as the ability to identify a sequence of steps necessary to solve a problem and to appreciate how future actions may alter circumstances (Lezak, 1995). Concretely, planning constitutes the second phase in the problem- solving framework of executive function, referring to the selection of the most efficient sequence of steps from many alternatives. The other three phases in sequence are representation, execution, and evaluation (Zelazo et al., 1997). Similarly, in the funnel model of executive function, which is proposed to account for learning disabilities and divides executive function into six processes, planning, denoted as the identification of short-term and long-term goals, is considered the first stage of problem solving, followed by organizing, prioritizing, shifting, memorizing, and checking (Meltzer, 2007, 2010; Meltzer & Krishnan, 2007). Also, planning, defined as the anticipation of future events, the formulation of a goal or endpoint, and the devise of actions to achieve goals, is one of the four components of Anderson’s (2002) executive system. The other three components are attentional control, cognitive flexibility, and information processing.

From a metacognitive perspective, planning is an important subcomponent of metacognition besides strategy selecting, monitoring, evaluating, and revising (Borkowski, Chan, & Muthukrishna, 2000), referring to the higher order mental process where individuals set short-term and long-term goals and form and select strategies to achieve these goals before learning (Pressley & Afflerbach, 1995; Schraw & Dennison, 1994; Schraw & Moshman, 1995). In the Cognitive-Metacognitive Framework for Studying Mathematical Performance, planning, together with orientation, execution, and verification, make up the four phases in performing a mathematical task (Garofalo & Lester, 1985). In the domain of writing, Zimmerman and Risemberg (1997) proposed three reciprocally interactive categories of self-regulatory processes that are used to regulate writing, including environmental, behavioral, and personal processes. Time planning and goal setting are two critical personal processes (Graham & Harris, 2000). Likewise, in Hayes and Flower’s (1980) and Bereiter and Scardamalia’s (1987) models of writing, planning is also considered one of the core components together with monitoring, evaluating, and revising that con- tribute to the accomplishment of writing tasks.

Sternberg (1984, 1985, 1987) made a distinction between meta- components, performance components, and knowledge-acquisition components with respect to cognitive ability. Metacomponents includes the executive skills, such as planning and monitoring, that people use to control their own information processing and lower order cognitive processes. Performance components are the lower order processes that are used to implement the commands of the metacomponents such as encoding and reasoning (Sternberg, 1984). It is evident that the definitions of planning in different fields are similar, and they all connote something at the metacognitive level. However, planning in the field of executive function and intelligence is measured by outcome-oriented performance on complex cognitive tasks, such as the Pattern Reasoning subtest of the KABC-II (A. S. Kaufman & Kaufman, 2004a) and the Matching Numbers subtest of the CAS (Naglieri & Das, 1997), as a performance component. Only planning in the field of metacognition is considered a metacomponent measured by questionnaires (Schraw & Dennison, 1994). What we derive from outcome- oriented measures is not a pure metacomponent but a combination of a metacomponent and a performance component. Accordingly, to distinguish between the two, the term *metacognitive planning* is used to refer to planning in the field of metacognition.

As previously mentioned, metacognitive planning has always been theoretically emphasized (Borkowski et al., 2000; Desoete, 2008; Flavell, 1979; Zimmerman, 1989). However, compared with other metacognitive components whose assessment methods are relatively more mature and supported by many empirical studies, metacognitive planning is less studied empirically due to the difficulty of separating it from task performance and other components. Though questionnaires can be easily administered and interpreted, they are vulnerable to social desirability (Richman, Kiesler, Weisband, & Drasgow, 1999), memory failure or distortion (Ericsson & Simon, 1993; Veenman, 2005), and interpretive reconstruction (Veenman, 2011a, 2011b). Therefore, there is a need for a reliable, valid, and pure method to assess metacognitive planning for both theoretical and practical purposes.

Assessment of Metacognitive Components

Though no valid tool specifically designed to measure metacognitive planning is currently available, we can gain some insights from assessments designed to measure other metacognitive components. A debate about how to assess metacognition has been ongoing for more than a decade (Baker & Cerro, 2000; Dinsmore, Alexander, &

Loughlin, 2008; Winne, 2010). Extant assessments of metacognition can be divided into offline and online methods (Veenman, 2005, 2011a, 2011b, 2013). *Offline methods* refer to questionnaires (*MAI; Schraw & Dennison, 1994) or interviews (Zimmerman & Martinez-Pons, 1990) used to collect data prior to or after tasks, and online methods are used to collect data during tasks, such as computer log-file analyses (Hadwin, Nesbit, Jamieson-Noel, Code, & Winne, 2007), observations (Whitebread et al., 2009), think-aloud protocols (Pressley & Afflerbach, 1995), and eye movement registration (Kinnunen & Vauras, 1995). Several studies have revealed that offline data does not correspond to actual behaviors (Hadwin et al., 2007; Winne & Jamieson-Noel, 2002) or to external criteria such as academic achievement (Desoete, Roeyers, & Buysse, 2001; Sperling, Howard, Staley, & DuBois, 2004), and online data are more closely related to learning outcomes (Veenman, 2005). Furthermore, correlations among offline methods are low to moderate, and correlations between online measures are moderate to high (Veenman, 2005). To sum up, offline methods are low in convergent and external validity com- pared with online methods, and accordingly, they may not accurately capture learners' metacognition. In this case, online methods are considered more appropriate.

Though more ecologically valid, online methods, such as observations and think-aloud protocols, must be administered individually and are, therefore, labor intensive and inconvenient for larger groups. Computer log-file analysis, which is used to collect data in a minimally intrusive way, can avoid the problems associated with observations and think-aloud protocols. *Computer log- file analysis* refers to the method of using trace data of cognitive tasks such as the number of mouth scroll, time spent on each stage, or moves taken to produce frequency count and patterns of study activities (Veenman, Bavelaar, De Wolf, & Van Haaren, 2014; Veenman, Wilhelm, & Beishuizen, 2004). However, there still exist some limitations of extant computer log-file analyses. First and foremost, an insufficient distinction is made between the underlying metacognitive components of each indicator. Rather, what we have is more like an umbrella term—*metacognitive skills* (Hadwin et al., 2007; Veenman et al., 2013). Actually, many indicators reflect a mixture of several metacognitive components (Veenman et al., 2013), making it hard to separate the various components and posing as an obstacle in the progress of both theory and practice. Second, criteria used for external validity have been limited to performances on specific tasks used to obtain the log-files (close criteria). Few studies up to now have examined the relationship between log-file data and more distal criteria, such as traditional educational outcome or work performance (Shraw, 2008), thus limiting the external validity. To sum up, the log-file analysis technique itself is not a concern, and we can obtain more valuable information if we choose purer indicators and broader criteria.

The Present Study

In the field of executive function and intelligence, planning, a process by definition, is measured by static performances on complex cognitive tasks. However, the results researchers derive from these measures are incongruent with the construct. In the field of metacognition, offline assessment of planning is low in validity. Though extant online methods are problematic in terms of either administration or measure clarity, they are still superior to offline methods in several ways. With advances in computer techniques, researchers can purposefully design tasks that require a student to externalize particular metacognitive activities (Baker & Cerro, 2000), as tracing these activities in computer log-files has become possible (Veenman et al., 2013; Winne, 2010). Taking advantage of computer log-file analysis, we aimed to develop in Study 1 a pure and standard measure of metacognitive planning, and Study 2 is dedicated to the systematic validation of its psychometric properties against multiple criteria.

Study 1: Preliminary Study

Method

Participants. The preliminary study went through two rounds of participant recruitment. In the first round, seven sophomores of psychology majors were recruited (three men, four women; M_{age} = 20.059, SD_{age} = 0.594). In the second round, 64 college students (19 men, 45 women; M_{age} = 20.735, SD_{age} = 1.702) from an optional course were recruited, and 10 credits were offered for their participation in the study.

Instrument development.

Selection of cognitive task. Metacognition must occur within a cognitive context (Li & Zhang, 2006). As the focus of this study is on metacognitive planning, researchers must design an appropriate cognitive task that necessitates metacognitive planning as a prerequisite for a successful solution. At the same time, the influence of extra variables, such as motivation and specific expertise, should be minimized or balanced as the task should be interesting enough to keep participants engaged and require little domain- specific expertise. On the basis of these principles and criteria, Sokoban, a Japanese puzzle game, was selected as the cognitive task. Sokoban consists of a pusher who must push a number of boxes into designated storage locations without getting the pusher, or box, stuck (as shown in Figure 1). The task is most often studied in artificial intelligence as an example of robot motion planning problems (Junghanns & Schaeffer, 2001). The pusher may only push one box at a time, cannot pull a box, and cannot occupy the same grid location as another box or barrier. A puzzle is considered solved when all boxes are pushed into the designated locations. The number of boxes and the number of locations are equal. Sokoban satisfies all of the requirements mentioned above as a candidate for measuring metacognitive planning, and its rules are quite simple.

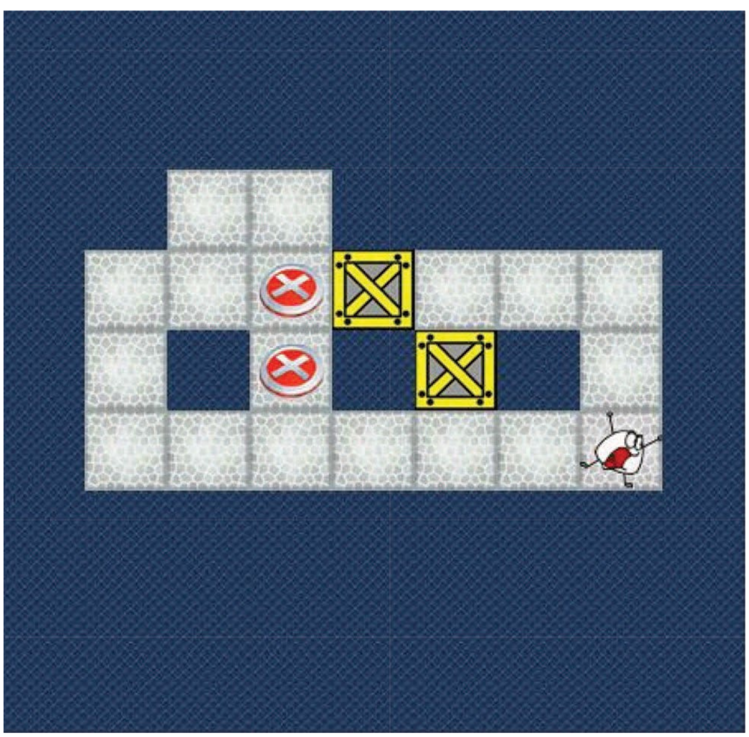

Figure 1. Example of Sokoban. See the online article for a color version of this figure.

Selection of measures. It is necessary to separate the meta- component from the performance component considering that this study is aimed at designing a pure measure of metacognitive planning. Because metacognition is a dynamic process (Nelson & Narens, 1990), a process-oriented index would better capture metacognitive planning than would be an outcome-oriented index. In the context of the Sokoban puzzle, there are at least two process-oriented indexes that researchers can use: (a) the absolute planning time and (b) the ratio between planning time and total time. As the former is thought to be prone to being mixed with processing speed or processing style, the authors of this study consider the latter index to be more appropriate. Empirical data are used to determine which index is better. As the Sokoban game requires planning ahead, people with higher levels of metacognitive planning are more likely to solve the puzzle in an effective manner. Therefore, performance indexes such as the number of successful solutions and extra moves can be used as validity criteria.

Item development. Researchers first searched for items with no more than four boxes on the Internet and, accordingly, retained 49 items. To induce metacognitive planning as much as possible and to best motivate participants, some items were redesigned so that the first move was crucial to the successful solution of all items, and the level of difficulty was maintained as moderate, on average, though with a wide range. To further maximize the importance of metacognitive planning, revocation of a move was not allowed, and there was only one attempt for each item. Finally, a pool with 36 items was formed.

The 36 items were written into a program using JAVA. Written instructions were presented on the right side of a 17-in. screen at all times. Demographic information, item number, number of total moves, planning time, total time, and solutions were recorded. Indexes such as metacognitive planning and extra moves can be derived from those recorded indexes. Details regarding these indexes are shown in Table 1.

Procedure. The 36 items were administered to the seven sophomores in a quiet classroom. After three practice trials, participants encountered the test phase in which they were instructed to be careful not to get stuck before initiating the first move. After the test, they were informed of the purpose of the study and were asked to provide feedback on any aspects of the game. On the basis of the data and the feedback, 11 items were excluded because of their high level of difficulty, and four items were modified to reduce their level of difficulty while maintaining the importance of the first move. Of the original items, 25 items were retained, rearranged by level of difficulty and then reprogrammed.

The 25 items were then individually administered to the second round of 64 participants in the psychology lab. The procedure was the same as that of the first administration. After finishing the assessment, participants were asked to rate the interestingness level of the game on a 3-point scale (1 = *very interesting*,3 = *very boring*).

Results and Discussion

The pass rate of each item was first calculated and equally distributed from 0.30 to 0.98. As the focus of this article is metacognitive planning instead of problem-solving ability, three items that no more than 50% of the college students were able to solve were excluded. As there were two pairs of items with the same difficulty level, the two items with larger number of optimal moves were excluded. In the end, 20 items with an equally distributed pass rate ranging from 0.52 to 0.98 were retained.

Descriptive and correlational statistics of the final 20 items are presented in Table 2. The average rating of interestingness regarding the Sokoban game was 1.422. Further analysis indicated that 43 participants (67.2%) considered the game very interesting, 15 participants (23.4%) considered the game moderately interesting, and only six participants (9.4%) considered the game very boring. Therefore, the Sokoban game was considered adequately interesting such that it could motivate participants to be sufficiently engaged and thereby bring out the participants' best efforts.

Cronbach's alphas of the absolute planning time, the time ratio, Sokoban score, and extra move were all above .75, indicating that these measures were reliable for further use. As for the two potential indexes of metacognitive planning, the time ratio was more internally consistent than was the absolute planning time (C = .938 and .874, respectively).

Table 1
Recorded and Derived Indexes (Study 1)

Index	Definitions
Recorded index	
Basic information	Name, gender, birthday, grade
Actual moves	The number of moves taken
Planning time	Time from the presence of each item to the first move of a box
Total time	Time from the presence of each item to the completion of each item
Solution	Whether participants solve each item; 1 = yes, 0 = no
Derived index	
Metacognitive planning	The ratio between planning time and total time
Extra moves	Actual number of moves minus optimal number of moves

With respect to the criterion validity, the time ratio accounted for 17.72% variance of the Sokoban score (the number of successful solution) and 38.07% variance of extra moves, and the absolute planning time explained 4.97% variance of the Sokoban score and 13.84% variance of extra moves. Though the time ratio and absolute planning time shared 63.68% common variance, it is evident that the time ratio was a better indicator of metacognitive planning in terms of both internal consistency and criterion validity.

Study 2

The time ratio was preliminarily found to be more appropriate. However, the sample size and the sample itself (all college students) were quite limited. Additionally, the criteria were quite restricted to indexes derived from Sokoban itself. They were insufficient to establish sound psychometric properties, and more distal criteria were needed. Therefore, Study 2 was conducted among four age groups with diverse criteria tests to systematically validate the time ratio.

Method

Participants. The participants in this study included 440 students sampled from four grades (fifth, seventh, and 10th grades and college) from four different schools in Beijing. Details about their age distribution are displayed in Table 3.

Measures. To validate the time ratio, researchers chose the Tower of London (ToL), intelligence test, and academic achievement as primary criteria. ToL is the most frequently used tool to assess planning ability in clinical and experimental neuropsychology (Albert & Steinberg, 2011; Berg & Byrd, 2002). Metacognitive components were also found to correlate moderately with intelligence (van der Stel & Veenman, 2008, 2010), and many learning theories have emphasized the role of metacognitive planning in academic achievement (Zimmerman, 1989). Academic motivation was used as a control variable when determining the unique contribution of metacognitive planning to academic achievement. Metacognitive questionnaires were also chosen to replicate previous studies that displayed low correlations with online measures (Veenman et al., 2013).

Because of practical constraints, not all measures were administered to every age group. Specifically, the Self-Report of Meta- cognitive Planning (SRMP), Sokoban, and Interestingness Rating (IR) were administered to all students; the Parent Rating of Meta- cognitive Planning (PRMP), the Tower of London (Kaller, Unterrainer, & Stahl, 2012), and the Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV; Zhang, 2008) were administered to 60 seventh-grade students; the Academic Motivation Scale (AMS; Vallerand et al., 1992) was administered to all the 10th- grade students.

Sokoban. The final 20 items were arranged by order of ascending difficulty. Written instructions are presented on the right side of the screen at all times. Three points are stressed in the instructions: (a) the basic rules of Sokoban, (b) be careful not to get stuck before initiating the first move because of no revocation and only one attempt for each item, (c) the need for quietness and independent work so as not to affect or be affected by others. When running the program, participants have to first fill in some basic information like birth date and gender. Three practice items are then presented to familiarize them with the game. They are required to successfully complete these three practice items before continuing. As for the test items, however, participants could not revoke their moves, and they are allowed only one attempt for each item.

Table 2
Descriptive and Correlational Statistics for Study Variables (N = 64; Study 1)

Measures	Cronbach α	*M*	*SD*	max	min	1	2	3	4
1. IN		1.422	0.622	3.000	1.000				
2. MP	.938	0.463	0.168	0.740	0.100	−.068			
3. SS	.753	16.031	3.212	20.000	8.000	−.350*	.421**		
4. EM	.758	11.238	5.686	27.150	1.050	.272*	−.617**	−.900**	
5. PT	.874	34.522	24.454	111.970	2.650	−.061	.798**	.223	−.372**

Note. IN = interestingness of Sokoban; MP = metacognitive planning; SS = Sokoban score; EM = extra moves; PT = planning time (s).
* $p < .05$. ** $p < .01$.

The ToL. Researchers chose the computerized version of ToL, which consisted of three pegs of different height and three balls of different colors— green, blue, and red. The psychometric proper- ties of the ToL have been validated (Kaller et al., 2012). Taking the younger age of our participants, eight 7-move items were excluded, leaving eight 4-move items, eight 5-move items, and eight 6-move items. Start state and goal state are presented on the upper and lower parts of the screen. Participants have to transform the start state into the goal state by moving the ball while following three rules: (a) one ball a time; (b) only the ball on the top can be moved; and (c) the longest peg can hold at most three balls, the second longest two balls at most, and the shortest only one ball at most. Any move that violates the rules will not be allowed. Whether a problem is solved in the minimum moves as well as all actual moves were recorded. For any further details, please refer to Kaller et al. (2012).

The WISC-IV. The Chinese version of the WISC-IV was used to assess participant intelligence. Ten core subtests, namely, Block Design, Picture Concepts, Matrix Reasoning, Similarities, Vocabulary, Comprehension, Digit Span, Letter-Number Sequencing, Coding, and Symbol Search, were administered to obtain four index scores and a full scale intelligence quotient (FSIQ). The reliability of these 10 subtests range from .78 to .92, and reliability of the FSIQ reached .97. Validity evidence has also been established (Zhang, 2008).

The SRMP. A scale with four items measuring metacognitive planning was adapted from Schraw and Dennison's (1994) MAI (e.g., "I set specific goals before I begin a task"). The items were translated into Chinese, back-translated into English and then translated into Chinese again by two researchers proficient in Chinese and English to avoid any potential misunderstandings and ensure translation accuracy. Some slight modifications in phrasing were made to make items more appropriate for students.

Participants were asked to rate statements on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Demographic information like birth date, gender, and parents' educational level are also included in this questionnaire.

The PRMP. Researchers adapted the subscale of the MAI to read from third-person perspective. For example, the original item, "I set specific goals before I begin a task" was rephrased as "He/She sets specific goals before he/she begins a task." Parents were instructed to rate their children on a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

The AMS. The 28-item AMS assesses students' motivation to attend high school, the psychometric properties of which have been tested (e.g., Fairchild, Horst, Finney, & Barron, 2005; Valle- rand et al., 1992). The AMS yields scores on three subscales of intrinsic motivation—to experience stimulation, to accomplish things, and to know; three subscales of extrinsic motivation— identified regulation, introjected regulation, and external regulation; and amotivation. Each subscale has four items. Participants used a 7-point Likert scale (from 1 = *does not correspond at all* to 7 = *corresponds exactly*). The translation procedure was the same as that for the SRMP. Cronbach's alpha of subscales in this study ranged from 0.75 to 0.86.

Academic achievement (AA). For the fifth-, seventh-, and 10th-grade students, final exam scores of two terms were used as indicators of AA. Scores for each term were first transformed into *z* scores and then averaged to obtain a single *z* score for each participants. A self-report grade-point average (GPA) procedure was adopted as academic indicators for college students.

Table 3
Sample Size and Age Distribution (Study 2)

	5th grade			7th grade			10th grade			College		
Variable	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total
N	37	44	81	60	53	113	59	68	127	21	98	119
M_{age}	11.770	11.405	11.572	12.713	12.617	12.669	16.048	15.903	15.971	21.115	20.670	20.749
SD	0.827	0.624	0.742	0.525	0.363	0.458	0.394	0.407	0.406	0.906	0.766	0.807

The IR. After completing Sokoban, one item ("How do you like the game") was used to assess the interestingness of the game on a 3-point scale (1 = *very interesting*, 3 = *very boring*).

Procedure. Sokoban was administered to college students individually and all other students by class. The ToL and WISC-IV were administered to 60 seventh graders individually. The PRMP were administered to parents of the seventh graders individually. All the other measures were administered by class.

Group administration. Metacognitive planning was assessed in a computer lab by class. An experienced experimenter and the supervisor of each class were present in the lab, the former responsible for possible questions from students or unexpected technical problems and the latter mainly responsible for class discipline. When participants were all seated and quiet, the experimenter explained the instructions to them. During the instruction phase, the experimenter was in control of the students' computers. When instructions were completed, students then ran the Sokoban program. After completion, the program terminated automatically. The experimenter then administered the questionnaire to the students. The entire process took, on average, 50 min. The AMS was administered to the 10th-grade students 2 months later in the classroom by class. An experienced experimenter first delivered instructions and then handed out the questionnaires. The supervisor of each class was also present for class discipline. It took 10 min on average to complete the AMS.

Individual administration. The WISC-IV and the ToL were administered to 60 students from two separate seventh-grade classes in the psychology lab. One class was composed of general population students, and the other was composed of talented students. Parents accompanied their children. Before the test, parents and their children read the consent form detailing the procedure, purposes, and potential problems of the test and then signed their names if both the parents and the child agreed to participate. The WISC-IV was administered in a quiet room by a trained experimenter in a standard way as recommended by the manual. After completing the WISC-IV, students were given a 10-min break before completing the ToL assessment. Instructions were identical to those of Kaller et al. (2012). The whole process took, on average, 140 min. While the child was completing the WISC-IV, the parents of the seventh graders were completing the PRMP in another room.

Retest. To test the stability of time ratio across different intervals and also for school constraints, retest was conducted 1 year later among the seventh graders, 4 months later among 10th graders, and 20 days later among college students. Retest data could not be obtained from fifth graders for some practical reasons. Specifically, 71 participants of seventh grade (34 females, M_{age} = 12.307, *SD* = 0.435), 48 participants of 10th grade (25 females, M_{age} = 15.466, *SD* = 0.287), and 32 college students (20 females, M_{age} = 20.811, *SD* = 0.813) took part in the retest. The procedure for retest was the same as the first test except that they did not have to complete questionnaires.

Results

Descriptive statistics for variables of interest are displayed in Table 4.

Interestingness. The average rating of interestingness of the Sokoban was 1.507 for the fifth graders, 1.664 for the seventh graders, 1.898 for the 10th graders, and 1.690 for college students. Further analysis revealed that 154 participants (36.2%) found it very interesting, 239 participants (56.2%) found it moderately interesting, and 32 participants (7.5%) found it very boring. Combined with data from Study 1, the Sokoban was again confirmed to be adequate to maintain participants' engagement. With the interestingness level as the independent variable and metacognitive planning as the dependent variable, a one-way analysis of variance (ANOVA) demonstrated that participant's perceived interesting- ness did not affect the time ratio score, $F(2, 438) = 0.240$, $p > .500$.

Reliability. The estimated internal consistency for time ratio scores ranged from .894 to .942 across the four age groups and remained .939 in the whole sample. Even the lower bound value was quite satisfactory. It was proposed that the alpha coefficient did not, per se, imply homogeneity (Cortina, 1993), and the precision of alpha in terms of the standard error of interitem correlation was therefore considered (Kaller et al., 2012). Mean interitem correlations were .440, and standard errors were .048.

Test–retest reliability was calculated using bivariate correlations between participants' time ratio scores on two tests. The test–retest correlations among the seventh-grade sample, 10th-grade sample, and college student sample were .582,

.649, and .786, respectively (*p*s $< .01$). Further, a paired sample *t* test demonstrated that the seventh graders' time ratio score on retest was significantly higher than that on the first test, $t(70) = 2.538$, $p = .013$, Cohen's $d = 0.276$ (corrected for the amount of correlation between paired measures based on Dunlap, Cortina, Vaslow, & Burke, 1996). Neither the 10th graders nor college students showed a significant difference between test and retest (*p*s $> .250$).

Split-half reliability was estimated by dividing the 20 items into two halves. Even items and odd items belonged to the two halves, respectively. The Spearman-Brown (S-B) coefficient for the time ratio score was .944.

As evidenced, the time ratio score showed good reliability, thus providing strong support for the further psychometric property analysis.

Unidimensionality. As all 20 items were intended to measure a single trait, a single-factor confirmatory factor analysis (CFA) was conducted to test the unidimensionality by maximum likelihood estimation with robust standard errors estimation for its tolerance of possible assumption violations (Muthén & Muthén, 2010). The single-factor model was found to fit the data well, S-Bx^2 (170) = 332.734, $p < .0001$, comparative fit index = .953, root-mean-square error of approximation = 0.047, $N = 440$. Therefore, the unidimensionality hypothesis was confirmed.

Developmental trend. A 4 (age group: fifth, seventh, 10th, college) X 2 (gender) between-subject ANOVA was conducted to explore the developmental trend and gender difference of meta- cognitive planning. The main effect of age was significant, $F(3, 432) = 45.481$, partial 'Y^2 = 0.24, $p < .0001$, whereas the main effect of gender and their interactions were not significant. A Bonferroni post hoc test revealed college students scored significantly higher than fifth- ($p < .001$) and seventh-grade students ($p < .001$); 10th-grade students scored significantly higher than fifth- ($p < .001$) and seventh-grade students ($p < .001$), and seventh-grade students scored significantly higher than fifth-grade students ($p = .018$). Though no significant difference was found, college students scored higher than 10th-grade students. Overall, the data showed an obvious developmental trend of metacognitive planning, as displayed in Figure 2, thus supporting the age sensitivity of the time ratio score.

Table 4
Descriptive Statistics for Study Variables (Study 2)

Variable	Statistic	IN	SRMP	MP	SS	EMS	PT
5th grade (n = 81)	Cronbach α		0.700	0.937	0.822	0.745	0.903
	M	1.507	4.460	0.291	7.099	24.071	9.356
	SD	0.578	0.622	0.174	4.215	4.738	9.501
7th grade (n = 113)	Cronbach α		0.621	0.942	0.848	0.814	0.897
	M	1.664	4.229	0.356	12.124	17.322	18.296
	SD	0.598	0.586	0.164	4.629	6.721	15.189
10th grade (n = 127)	Cronbach α		0.589	0.895	0.753	0.737	0.830
	M	1.898	4.111	0.482	16.024	11.061	28.196
	SD	0.602	0.662	0.130	3.216	5.441	16.439
College (n = 119)	Cronbach α		0.623	0.894	0.797	0.770	0.868
	M	1.69	3.865	0.508	16.059	10.561	35.638
	SD	0.55	0.530	0.127	3.477	5.869	21.471

Note. IN = interestingness of Sokoban; SRMP = Self-Report of Metacognitive Planning; MP = metacognitive planning; SS = Sokoban score; EMS = extra moves of sokoban; PT = planning time (s).

Criteria validity. As mentioned in the introduction, criteria is classified as close and distal. In this study, criteria obtained from the Sokoban game were defined as close criteria, those obtained from the ToL were defined as moderately close criteria, and AA as well as self/parent rating were defined as distal criteria. In the present study, close criteria and distal criteria were obtained from all age groups, and moderately close criteria were obtained from 60 seventh-grade students. Significant but decreasing correlations were expected as the criteria became more distal. Descriptive and correlational statistics are shown in Table 5.

Close criteria. Extra moves and successful solutions of Sokoban were used as close criteria. The time ratio was positively correlated with them ($r_{Extra\text{-}MP}$ = -.705, $r_{Solution\text{-}MP}$ = .577, *p*s <.01). Even after controlling for participant motivation to play the game, their self/parents rating and intelligence, metacognitive planning still accounted for 36.1% and 28.8% variance of extra moves and successful solutions, respectively.

Moderately close criteria. Extra moves and successful solutions of ToL were treated as moderately close criteria. The time ratio was positively correlated with them ($r_{Extra\text{-}MP}$ = -.301, $r_{Solution\text{-}MP}$ = .562, *p*s < .0001). Metacognitive planning explained 6.1% and 15.2% variance of extra moves and successful solutions, respectively, after controlling for self/parents rating and intelligence in the seventh grade.

Distal criteria. As for the 10th graders, results suggested that metacognitive planning correlated with mathematics scores, physics scores, chemistry scores, history scores, and geology scores (r = .294, .378, .290, .232, .292, respectively; ps < .01). As thoroughly studied, AA was affected by many factors, and motivation was one of the most important (Zimmerman, Bandura, & Martinez-Pons, 1992). Hierarchical regressions were conducted to examine the unique contribution of metacognitive planning to AA. Gender and parents' educational level were entered first, followed by academic motivation second and metacognitive planning. The results showed that even after controlling for these variables, metacognitive planning still accounted for 7.7%, 12.2%, 6.8%, 4.8%, and 9.4% of the variance in the mathematics score, physics score, chemistry score, history score, and geology score, respectively. Metacognitive planning had no significant effect on Chinese and English language scores. Details of the regression results are shown in Table 6.

For the fifth- and seventh-grade students, metacognitive planning could not predict their AA in any of the subjects. Also, no significant correlation was found between metacognitive planning and GPA. As is expanded on in the Discussion, different reasons may underlie the two seemingly same results; namely, metacognitive planning was predictive neither of AA of college students nor of the younger students (the fifth and seventh graders).

As for self-rating and parent rating, there was a significant but small correlation between the time ratio and the self-rating (r =.179, p < .01), and no significant correlation was found between the time ratio and parent rating.

In the seventh-grade sample, an independent sample t test con- firmed the intelligence difference between the two classes, $t(59) = 6.423$, Cohen's d = 1.672, p < .0001. Another t test revealed that the talented class demonstrated a significantly higher level of metacognitive planning, $t(59) = 4.989$, Cohen's d = 1.299, p < .0001. Further correlation analysis revealed a significant correlation between intelligence and metacognitive planning (r = .470, p < .0001).

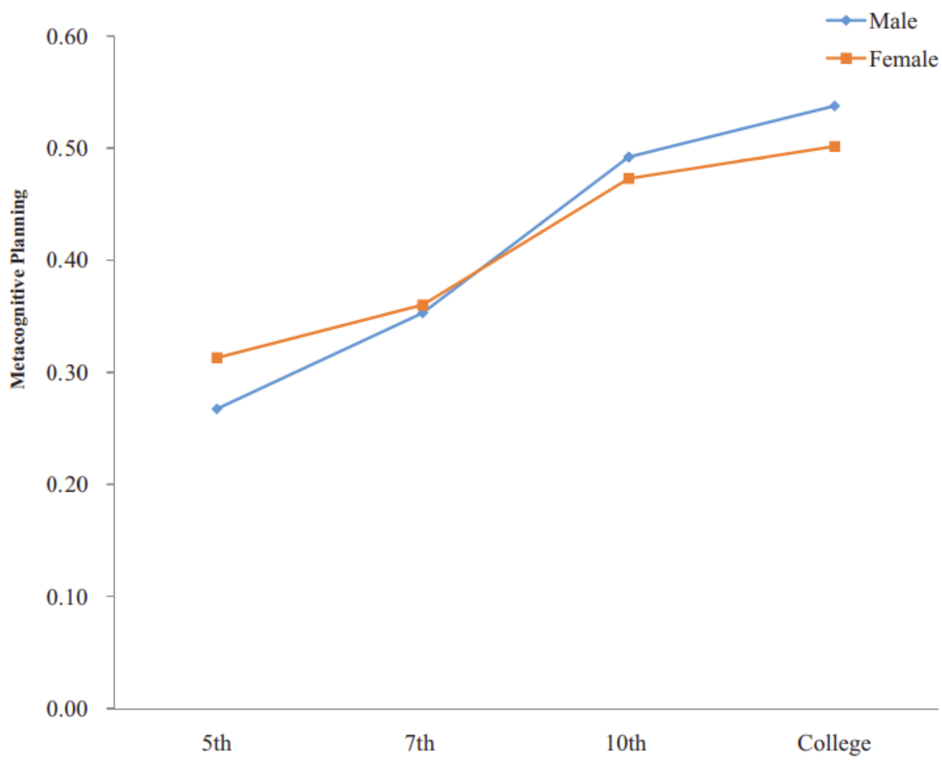

Figure 2.

Developmental trend of metacognitive planning. See the online article for a color version of this figure.

Table 5

Descriptive and Correlational Statistics for Study Variables (Study 2)

Measure	*N*	Cronbach's α	*M*	*SD*	1	2	3	4	5	6	7	8
1. PRMP	60	.679	4.021	0.505								
2. MP	440	.939	0.422	0.170	.059							
3. FSIQ	60	.970	148.661	15.295	.094	.470**						
4. ToLS	60	.589	16.508	2.932	.016	.562**	.574**					
5. EMT	60	.479	1.034	0.504	−.122	−.310*	−.306*	−.670**				
6. SS	440	.892	13.389	5.155	−.226	.520**	.483**	.616**	−.281*			
7. EMS	440	.864	14.929	7.731	.106	−.604**	−.490**	−.649**	.405**	−.955**		
8. PT	440	.894	24.198	19.114	.076	.786**	.354**	.435**	−.221	.786**	−.536**	
9. SRMP	419	.656	4.138	0.633	.142	.179**	.315*	.308*	−.094	.180**	−.208**	.147**

Note. PRMP = Parent Rating of Metacognitive Planning; MP = metacognitive planning; FSIQ = full scale intelligence quotient; ToLS = Tower of London (ToL) score; EMT = extra move of ToL; SS = Sokoban score; EMS = extra moves of Sokoban; PT = planning time (s); SRMP = Self-Report of Metacognitive Planning.

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

Discussion

Given the importance of metacognitive planning in many fields and the shortage of psychometrically sound instruments, this re- search focused on the measurement of metacognitive planning as a scientific priority. Increased precision of measurement will facilitate a more thorough understanding of the developmental course and its different contributions to different areas. With these goals in mind, this study reported the psychometric properties, including reliability, structural validity, and criteria validity of a newly developed task designed to measure metacognitive planning in a sample consisting of four different age groups.

No more than 10% of the participants considered the Sokoban game boring in either the preliminary sample or the formal sample (9.4% and 7.5%, respectively). Thus, it can be concluded that the Sokoban game is an appropriate choice with respect to participant engagement. Whether participants perceived Sokoban interesting or not, their time ratio scores were unaffected, thus demonstrating that the time ratio was a stable measure across different conditions.

High Cronbach's alpha, high split-half reliability, and low standard error of interitem correlations together demonstrate that the time ratio is a rather pure measure. Further, CFA again confirmed the unidimensionality of the time ratio. Also, testetest reliabilities in the college group is quite satisfactory. Though lower than college group, test–retest reliabilities in both the seventh-grade group and the 10th-grade group are also quite acceptable considering the long intervals and rapid development in this age period. In sum, the time ratio is both a pure and temporally stable measure.

Consistent with previous studies on metacognition development (Baker & Brown, 1984; Hacker, Dunlosky, & Graesser, 1998; McCormick, 2003), the time ratio also revealed an obvious developmental trend of metacognitive planning. Veenman et al. (2004) obtained a monotonic maturation effect of metacognitive skills across age groups (nine, 12, 14, and 22) using four different discovery-learning tasks. Due to the limitations of discovery- learning tasks, a study to determine whether the monotonic maturation hypothesis held in problem-solving tasks was conducted, and results again confirmed the monotonic maturation hypothesis (Veenman & Spaans, 2005). A similar growth in both quantity and quality of metacognitive skills, especially planning and evaluating, was found by Van der Stel, Veenman, Deelen, and Haenen (2010). The results from longitudinal studies further confirmed the mono- tonic maturation effect of both quantity and quality of metacognitive skills, especially planning (van der Stel & Veenman, 2010). Researchers have demonstrated that the greatest change in meta- cognitive skills occurred between 10 and 14 years of age (Bereiter &

Scardamalia, 1987). Our measure strongly validates these results.

Many researchers have delved into the relationship between metacognition and intelligence and reached, to some degree, consensus that metacognitive skills and intelligence are two dependent (r = .45, on average) but distinct constructs (Helms-Lorenz & Jacobse, 2008; Veenman et al., 2004; Veenman & Spaans, 2005; van der Stel, & Veenman, 2008, 2010). Similarly, our measure revealed a comparable correlation between metacognitive planning and intelligence, thus contributing to the validity of inference made from the time ratio score.

The significant but small correlation between the time ratio and self-report is similar to what Veenman et al. (2013) reported (r = .18 in the present study and r = .15 in Veenman's study). And the correlation between the time ratio and successful solutions of the ToL is much higher (r = .562). These results are consistent with previous studies that revealed that although online measures converged with each other (Cromley & Azevedo, 2006), they did not converge with offline measures (Bannert & Mengelkamp, 2008). Therefore, interpretations made from the time ratio is more valid than that made from questionnaires.

The fact that metacognitive planning showed significant correlations with Sokoban performance and ToL performance can be explained by the nature of the two tasks as both tasks rely heavily on planning (Berg & Byrd, 2002; Junghanns & Schaeffer, 2001). Moreover, because successful solution is the main goal for participants, metacognitive planning accounted for a larger proportion of successful solution variance than it did for extra moves. Due to the restricted range of ToL extra moves, a smaller correlation was expected than that of Sokoban extra moves. Furthermore, the relatively higher correlation with close criteria than that with moderately close criteria is because both the time ratio and Sokoban performance are derived from the same task. Task specificity contributes to this small difference. In sum, both close and moderately close criteria provide strong validity evidence for our measure.

Though the distal criteria complicate the matter, they still sup- port the validity of inference made from the time ratio score. First, the relatively smaller correlation compared with that of close and moderately close criteria is due to the complexity of AA. AA is affected by many factors, such as intelligence (Deary, Strand, Smith, &

Fernandes, 2007), motivation (Zimmerman & Schunk, 2001), socioeconomic status (Sirin, 2005), personality (Poropat, 2009), and metacognition (Paris & Winograd, 1990; Schunk, 2008). Therefore, the remaining explanatory power of metacogni- tive planning after statistical control is actually quite satisfactory. Second, the results demonstrate that metacognitive planning is only predictive of 10th-grade students' AA, a finding that is partially consistent with previous studies. Both cross-sectional and longitudinal studies have revealed that metacognitive skills were more predictive of 15-year-old students' math score than that of the 13-year-old students (van der Stel et al., 2010; van der Stel & Veenman, 2010). Our measure strengthens and broadens this con- clusion. The different contributions of metacognitive planning to AA could be explained from two perspectives. From a develop- mental view, though there is ample evidence that preschool chil- dren demonstrate metacognitive skills such as monitoring and planning with certain tasks (Balcomb & Gerken, 2008; Lyons & Ghetti, 2013; Paulus, Proust, & Sodian, 2013), these skills develop at an implicit level (Paulus et al., 2013) and become more explicit and sophisticated only after they go to school and receive challenging courses (Veenman, Van Hout-Wolters, & Afflerbach, 2006). It is generally assumed that explicit metacognitive skills such as planning emerge at the age of 8 –10 (Berk, 2006). From the domain-general/ domain-specific perspective, Veenman and Spaans (2005) proposed that metacognitive skills originate as domain specific in early child- hood and develop to be domain general at the age of 12. Accordingly, the metacognitive planning for participants in the fifth and seventh grades are still developing. Therefore, they are still too premature and domain specific (Veenman & Spaans, 2005) to enhance academic-related activities. With respect to college students, the GPA may not be a good indicator of AA, as different teachers enforce different standards (Goldman & Slaughter, 1976). Also, the reliability and validity of self-reported GPA were questionable (Kuncel, Credé, & Thomas, 2005). Furthermore, in mainland China, there are several heavily weighted but less meaningful compulsory courses that affect the validity of the GPA. Therefore, the insignificant correlation may be attributed to the problematic reliability and validity of the criteria. Overall, the distal criteria also support the validity of inferences made from the time ratio score.

Table 6

Regression of Academic Achievement on Metacognitive Planning (10th Grade; Study 2)

Variable	Chinese		English		History		Math		Physics		Chemistry		Geology	
	β	*t*	β	*t*	β	*t*	β	*t*	β	*t*	β	*t*	β	*t*
1. Demographics (DE)														
GE	−0.138	−1.463	−0.215**	−2.448**	−0.032	−0.338	0.189*	2.149*	0.188*	2.170*	0.143	1.597	−0.194*	−2.198*
ME	0.058	0.560	0.155	1.594	0.014	0.138	0.145	1.483	0.029	0.302	0.123	1.243	−0.023	−0.238
FE	−0.028	−0.270	0.086	0.877	−0.045	−0.433	−0.028	−0.287	0.035	0.358	−0.039	−0.386	0.002	0.019
2. Motivation (MO)														
IM	0.049	0.437	0.108	1.027	0.196	1.753	0.119	1.130	0.137	1.322	0.212**	1.983*	0.300**	2.850**
EM	−0.096	−0.946	−0.167	−1.760	−0.113	−1.111	0.108	1.131	0.014	0.151	−0.062	−0.637	−0.191*	−2.001*
AM	−0.134	−1.310	−0.197**	−2.051*	−0.029	−0.283	−0.055	−0.577	−0.086	−0.910	−0.079	−0.805	0.077	0.796
3. Metacognition (ME)														
MP	0.183	1.976	0.136	1.570	0.224*	2.433*	0.284**	3.274**	0.357**	4.200**	0.268**	3.087**	0.314**	3.616**
	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2	Adj. R^2	ΔR^2
Model														
1. DE	−0.006	0.020	0.082	0.105**	−0.025	0.001	0.029	0.053	0.019	0.043*	0.005	0.030	0.006	0.031
2. DE + MO	0.002	0.033	0.139	0.077*	−0.005	0.044	0.075	0.068*	0.058	0.062	0.055	0.073*	0.054	0.071*
3. DE + MO + MP	0.027	0.032	0.150	0.018	0.036	0.048*	0.148	0.077**	0.179	0.122**	0.119	0.068**	0.150	0.094**

Note. GE = gender; ME = mother education; FE = father education; IM = intrinsic motivation; EM = extrinsic motivation; AM = amotivation; MP = metacognitive planning.

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

Although some researchers have been developing online assessments of metacognition with the emergence of the computer-based learning environment (Kunz, Drewniak, & Schott, 1992; Veenman et al., 2004; Veenman et al., 2013; Winne, 2010), most of the studies suffer from several limitations. First, these measures tend to yield a mixture of metacognitive components (Veenman et al., 2013), thus offering a limited contribution to a thorough under- standing of the development or the underlying mechanism of how each metacognitive component helps people in real life. If further intervention is intended, a clear understanding of the interactive function of different metacognitive components is essential. Second, as many studies that focused on online assessments have not been conducted from a psychometric perspective (e.g., Hadwin et al., 2007; Veenman et al., 2004, 2005), they have been inadequate in clarifying the type of log-file, the reliability of the indicator, and the unidimensionality of the adopted index. Though criterion validity is generally reported, the criteria are quite limited to close criteria. Less moderately close and few distal criteria have been adopted, which restrict their validity. More seriously, the research samples are usually quite small and nonrepresentative. Accordingly, the psychometric evidence of these measures is insufficient.

With respect to this study, it was conducted mainly from a psychometric perspective. We centered on a single important component of metacognition, used a larger sample, and were clear of what the log-file represented. The evidence for reliability was robust, and an obvious developmental trend consistent with numerous previous studies provided solid validity support. We innovatively classified criteria into three levels, and results again confirmed our hypothesis that there would be significant but decreasing correlations as the criteria became more distal, which further strengthened the validity of inference that could be made from the measure. Taken together, those lines of evidence demonstrate that this measure is psychometrically sound.

There are, however, at least two limitations to which future studies must pay attention. First, participants of all ability levels were administered the same items. Accordingly, there is likely to be some degree of mismatch between participants and items, which may result in inaccurate estimates of their metacognitive planning. If possible, computerized adaptive testing techniques (Wainer et al., 1990) should be used to select the most suitable items for each participant, thereby yielding more accurate estimates. Second, the psychometric properties of the AA indicators in this study are actually unknown. Unlike the Western world where there are various types of standardized achievement tests, such as the Wide Range Achievement Test, Fourth Edition (Wilkinson & Robertson, 2006), the Wechsler Individual Achievement Test, Third Edition (Wechsler, 2009), and the Kaufman Tests of Educational Achievement, Second Edition (A. S. Kaufman & Kaufman, 2004b), there are few standardized AA tests available in China. Therefore, an academic indicator with known psychometric properties would further clarify the intricate relationship between metacognitive planning and academic achievement.

Though many theorists emphasize the importance of metacognitive planning, empirical support for these theories depends on tasks that can precisely capture it. For further insight, it is imperative to develop a scalable instrument to facilitate inferences about individual differences, developmental courses, and differential contributions to different areas of metacognitive planning. Taking advantage of computer techniques, this study represents our unremitting efforts toward achieving this goal.

References

Albert, D., & Steinberg, L. (2011). Age differences in strategic planning as indexed by the Tower of London. *Child Development, 82,* 1501–1517. doi:10.1111/j.1467-8624.2011.01613.x

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology, 8,* 71– 82. doi:10.1076/ chin.8.2.71.8724

Baker, L., & Brown, A. L. (1984). Metacognitive skills and reading. In P. D. Pearson, R. Barr, & P. Mosenthal (Eds.), *Handbook of research in reading* (Vol. 1, pp. 353–395). New York, NY: Longman.

Baker, L., & Cerro, L. C. (2000). Assessing metacognition in children and adults. In J. C. Impara, L. L. Murphy, & G. Shraw (Eds.), *Issues in the measurement of metacognition* (pp. 99 –146). Lincoln, NE: Buros Institute of Mental Measurements.

Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science, 11,* 750 –760. doi:10.1111/j.1467-7687.2008.00725.x

Bannert, M., & Mengelkamp, C. (2008). Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning? *Metacognition and Learning, 3,* 39 –58. doi:10.1007/s11409-007-9009-6

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written com- position*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Berg, W. K., & Byrd, D. L. (2002). The Tower of London spatial problem- solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology, 24,* 586 – 604. doi: 10.1076/jcen.24.5.586.1006

Berk, L. (2006). *Child development* (7th ed.). Boston, MA: Allyn & Bacon.

Borkowski, J. G., Chan, L. K., & Muthukrishna, N. (2000). A process-oriented model of metacognition: Links between motivation and executive functioning. In J. C. Impara, L. L. Murphy, & G. Shraw (Eds.), *Issues in the measurement of metacognition* (pp. 1– 42). Lincoln, NE: Buros Institute of Mental Measurements.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78,* 98 –104. doi: 10.1037/0021-9010.78.1.98

Cromley, J. G., & Azevedo, R. (2006). Self-report of reading comprehension strategies: What are we measuring? *Metacognition and Learning, 1,* 229 –247. doi:10.1007/s11409-006-9002-5

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence*. Needham Heights, MA: Allyn & Bacon.

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

Desoete, A. (2008). Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning, 3,* 189 –206. doi:10.1007/s11409-008-9026-0

Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34,* 435– 447.

doi:10.1177/002221940103400505

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20,* 391– 409. doi:10.1007/ s10648-008-9083-6

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 1,* 170 –177. doi:10.1037/1082-989X.1.2.170

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, MA: MIT Press.

Fairchild, A. J., Horst, S. J., Finney, S. J., & Barron, K. E. (2005). Evaluating existing and new validity evidence for the Academic Motivation Scale. *Contemporary Educational Psychology, 30,* 331–358. doi: 10.1016/j.cedpsych.2004.11.001

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychologist, 34,* 906 – 911. doi:10.1037/0003-066X.34.10.906

Garden, S. E., Phillips, L. H., & MacPherson, S. E. (2001). Midlife aging, open-ended planning, and laboratory measures of executive function. *Neuropsychology, 15,* 472– 482. doi:10.1037/0894-4105.15.4.472

Garofalo, J., & Lester, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16,* 163–176. doi:10.2307/748391

Goldman, R. D., & Slaughter, R. E. (1976). Why college grade point average is difficult to predict. *Journal of Educational Psychology, 68,* 9 –14. doi:10.1037/0022-0663.68.1.9

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35,* 3–12. doi:10.1207/S15326985EP3501_2

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.). (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Meta- cognition and Learning,* *2,* 107–124. doi:10.1007/s11409-007-9016-7

Hayes, J. R., & Flower, L. S. (1980). Identifying the organization of the writing process. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3–30). Hillsdale, NJ: Erlbaum.

Helms-Lorenz, M., & Jacobse, A. E. (2008). Metacognitive skills of the gifted from a cross-cultural perspective. In M. F. Shaughnessy, M. V. Veenman, & C. K. Kennedy (Eds.), *Metacognition: A recent review of research, theory, and perspectives* (pp. 3– 43). Happauge, NY: Nova.

Junghanns, A., & Schaeffer, J. (2001). Sokoban: Enhancing general single-agent search methods using domain knowledge. *Artificial Intelligence, 129,* 219 –251.

Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment, 24,* 46 –53. doi:10.1037/a0025174

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Assessment Battery for Children–II (KABC-II)*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, J. C., Kaufman, A. S., Kaufman-Singer, J., & Kaufman, N. L. (2005). The Kaufman Assessment Battery for Children–Second Edition. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed., pp. 344 –370). New York, NY: Guilford Press.

Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction, 5,* 143–165. doi:10.1016/ 0959-4752(95)00009-R

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta- analysis and review of the literature. *Review of Educational Research, 75,* 63– 82. doi:10.3102/00346543075001063

Kunz, G. C., Drewniak, U., & Schott, F. (1992). On-line and off-line assessment of self-regulation in learning from instructional text. *Learning and Instruction, 2,* 287–301. doi:10.1016/0959-4752(92)90019-I

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York, NY: Oxford University Press.

Li, J., & Zhang, H. C. (2006). Characteristics of on-line metacognitive regulating ability under cognitive context [in Chinese]. *Acta Psychologica Sinica, 38,* 342–348.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222,* 66 –78. doi:10.1038/scientificamerican0370-66

Lyons, K. E., & Ghetti, S. (2013). I don't want to pick! Introspection on uncertainty supports early strategic behavior. *Child Development, 84,* 726 –736. doi:10.1111/cdev.12004

McCormick, C. B. (2003). Metacognition and learning. In W. M. Reynolds & G. E. Miller (Eds.), *Handbook of psychology: Educational psychology* (pp. 79 –102). Hoboken, NJ: Wiley. doi:10.1002/0471264385.wei0705 Meltzer, L. (Ed.). (2007). *Executive function in education: From theory to practice*. New York, NY: Guilford Press.

Meltzer, L. (Ed.). (2010). *Promoting executive function in the classroom*. New York, NY: Guilford Press.

Meltzer, L., & Krishnan, K. (2007). Executive function difficulties and learning disabilities: Understandings and misunderstandings. In L. Meltzer (Ed.), *Executive function in education: From theory to practice* (pp. 77–105). New York, NY: Guilford Press.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide, v 6.1*. Los Angeles, CA: Muthén & Muthén.

Naglieri, J. A. (2011). PASS theory. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of child behavior and development* (p. 1069). New York, NY: Springer.

Naglieri, J. A., & Das, J. P. (1988). Planning-arousal-simultaneous- successive (PASS): A model for assessment. *Journal of School Psychology, 26,* 35– 48. doi:10.1016/0022-4405(88)90030-1

Naglieri, J. A., & Das, J. P. (1990). Planning, attention, simultaneous, and successive (PASS) cognitive processes as a model for intelligence. *Journal of Psychoeducational Assessment, 8,* 303–337. doi:10.1177/ 073428299000800308

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. Itasca, IL: Riverside.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125–173). San Diego, CA: Academic Press. doi:10.1016/S0079-7421(08)60053-5

Paris, S. G., & Winograd, P. (1990). How metacognition can promote academic learning and instruction. In B. F. Jones & L. Idol (Eds.), *Dimensions of thinking and cognitive instruction* (pp. 15–51). Hillsdale, NJ: Erlbaum.

Paulus, M., Proust, J., & Sodian, B. (2013). Examining implicit metacognition in 3.5-year-old children: An eye-tracking and pupillometric study. *Frontiers in Psychology, 4,* 1–7. doi:10.3389/fpsyg.2013.00145

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin, 135,* 322–338. doi:10.1037/a0014996

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale, NJ: Erlbaum.

Richman, W. L., Kiesler, S., Weisband, S., & Drasgow, F. (1999). A meta-analytic study of social desirability distortion in computer- administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology, 84,* 754–775. doi:10.1037/0021-9010.84.5.754

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19,* 460–475. doi:10.1006/ceps.1994.1033

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7,* 351–371. doi:10.1007/BF02212307

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review, 20,* 463– 467. doi:10.1007/s10648-008-9086-3

Shraw, G. (2008). Measuring metacognitive judgments. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 415–429). Madison, NY: Routledge.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417– 453. doi:10.3102/00346543075003417

Sperling, R. A., Howard, B. C., Staley, R., & DuBois, N. (2004). Meta- cognition and self-regulated learning constructs. *Educational Research and Evaluation, 10,* 117–139. doi:10.1076/edre.10.2.117.27905

Sternberg, R. J. (1984). Toward a triarchic theory of human intelligence. *Behavioral and Brain Sciences, 7,* 269 –316. doi:10.1017/ S0140525X00044629

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York, NY: Cambridge University Press.

Sternberg, R. J. (1987). Synopsis of a triarchic theory of human intelligence. In S. H. Irvine & S. E. Newsstead (Eds.), *Intelligence and cognition: Contemporary frames of reference* (pp. 141–176). Dordrecht, the Netherlands: Kluwer Academic.

Vallerand, R. J., Pelletier, L. G., Blais, M. R., Briere, N. M., Senecal, C., & Vallieres, E. F. (1992). The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educational and Psychological Measurement, 52,* 1003–1017. doi:10.1177/0013164492052004025

van der Stel, M., & Veenman, M. V. J. (2008). Relation between intellectual ability and metacognitive skillfulness as predictors of learning performance of young students performing tasks in different domains. *Learning and Individual Differences, 18,* 128 –134. doi:10.1016/j.lindif.2007.08.003

van der Stel, M., & Veenman, M. V. J. (2010). Development of metacognitive skillfulness: A longitudinal study. *Learning and Individual Differences, 20,* 220 –224. doi:10.1016/j.lindif.2009.11.005

Van der Stel, M., Veenman, M. V., Deelen, K., & Haenen, J. (2010). The increasing role of metacognitive skills in math: A cross-sectional study from a developmental perspective. *ZDM, 42,* 219 –229. doi:10.1007/ s11858-009-0224-2

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* [Learning strategies and metacognition: Implications for research and practice] (pp. 77–99). Münster, Germany: Waxmann.

Veenman, M. V. (2011a). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6,* 205–211. doi:10.1007/s11409-011-9080-x

Veenman, M. V. J. (2011b). Learning to self-monitor and self-regulate. In R. Mayer & P. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 197–218). New York, NY: Routledge.

Veenman, M. V. J. (2013). Assessing metacognitive skills in computerized learning environments. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies* (pp. 157–168). New York, NY: Springer. doi:10.1007/978-1-4419-5546-3_11

Veenman, M. V. J., Bavelaar, L., De Wolf, L., & Van Haaren, M. G. P. (2014). The on-line assessment of metacognitive skills in a computerized learning environment. *Learning and Individual Differences, 29,* 123–130. doi:10.1016/j.lindif.2013.01.003

Veenman, M. V. J., & Spaans, M. A. (2005). Relation between intellectual and metacognitive skills: Age and task differences. *Learning and Indi- vidual Differences, 15,* 159 –176. doi:10.1016/j.lindif.2004.12.001

Veenman, M. V. J., Van Hout-Wolters, B. H., & Afflerbach, P. (2006). Metacognition and learning: Conceptual and methodological considerations. *Metacognition and Learning, 1,* 3–14. doi:10.1007/s11409-006- 6893-0

Veenman, M. V. J., Wilhelm, P., & Beishuizen, J. J. (2004). The relation between intellectual and metacognitive skills from a developmental perspective. *Learning and Instruction, 14,* 89 –109. doi:10.1016/j.learninstruc.2003.10.004

Wainer, H., Dorans, N. J., Green, B. F., Steinberg, L., Flaugher, R., Mislevy, R. J., & Thissen, D. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Erlbaum.

Wechsler, D. (2009). *Wechsler Individual Achievement Test—third edition*. San Antonio, TX: Psychological Corporation.

Whitebread, D., Coltman, P., Pasternak, D. P., Sangster, C., Grau, V., Bingham, S., & Demetriou, D. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4,* 63– 85. doi:10.1007/ s11409-008-9033-1

Wilkinson, G. S., & Robertson, G. J. (2006). *Wide Range Achievement Test (WRAT4)*. Lutz, Germany: Psychological Assessment Resources.

Willcutt, E. E., Doyle, A. E., Nigg, J. T., Faraone, S. S., & Pennington,B. F.

(2005). Validity of the executive function theory of attention- deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57,* 1336 –1346. doi:10.1016/j.biopsych.2005.02.006

Winne, P. H. (2010). Improving measurements of self-regulated learning. *Educational Psychologist, 45,* 267–276. doi:10.1080/00461520.2010.517150

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology, 27,* 551–572. doi:10.1016/S0361-476X(02)00006-1

Zelazo, P. D., Carter, A., & Reznick, J. S. (1997). Early development of executive function: A problem-solving framework. *Review of General Psychology, 1,* 198 –226. doi:10.1037/1089-2680.1.2.198

Zhang, H. C. (2008). *Chinese version of the Wechsler Intelligence Scale for Children–Fourth Edition*. Zhuhai, Guangzhou: King-May Psychological Assessment.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology, 81,* 329 –339. doi: 10.1037/0022-0663.81.3.329

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self- motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal, 29,* 663– 676. doi:10.3102/00028312029003663

Zimmerman, B. J., & Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self- efficacy and strategy use. *Journal of Educational Psychology, 82,* 51–59. doi:10.1037/0022-0663.82.1.51

Zimmerman, B. J., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22,* 73–101. doi:10.1006/ceps.1997.0919

Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Mahwah, NJ: Routledge.