

## The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving

Donna Bryce, David Whitebread

### Angaben zur Veröffentlichung / Publication details:

Bryce, Donna, and David Whitebread. 2012. "The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving." *Metacognition and Learning* 7 (3): 197–217.  
<https://doi.org/10.1007/s11409-012-9091-2>.



# The development of metacognitive skills: evidence from observational analysis of young children's behavior during problem-solving

Donna Bryce · David Whitebread

**Abstract** This study aimed to better understand how metacognitive skills develop in young children aged 5 to 7 years. In particular, we addressed whether developmental changes reflect quantitative or qualitative improvements, and how metacognitive skills change with age and task-specific ability. Previous research into the development of metacognitive skills has been somewhat limited by methodology—often there is an over-reliance on language skills and it is assumed that children are fully conscious of the skills they use. In this study, a new observational method was developed which aimed to better represent young children's ( $n=66$ ) metacognitive skills by coding their verbalizations and non-verbal behavior during a problem-solving task. This method proved to be developmentally sensitive and illustrated both a quantitative increase in metacognitive skills, and qualitative changes in the types of monitoring and planning used throughout early development. Further, the results indicated that monitoring processes improve with age, control processes improve with both age and task-specific ability, and 'failures of metacognitive skills' are primarily affected by task-specific ability rather than age.

**Keywords** Metacognitive skills · Monitoring · Control · Planning · Problem-solving · Children · Development · Observational coding

---

This work was funded by an ESRC PhD studentship grant awarded to D.B. Thank you to Dr Sue Bingham and Charlotte Bargas for second coding.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11409-012-9091-2) contains supplementary material, which is available to authorized users.

D. Bryce · D. Whitebread  
Faculty of Education, University of Cambridge, 184 Hills Road, Cambridge, CB2 8PQ, UK

D. Bryce (✉)  
Fachbereich Psychologie, Universität Tübingen, Schleichstrasse 4, 72072 Tübingen, Germany  
e-mail: donna.bryce@uni-tuebingen.de

## Introduction

Understanding how typically developing children acquire the ability to monitor and control their cognition on-task (metacognitive skills) has to date been restricted by methodological challenges (De la Fuente and Lozano 2010; Winne and Perry 2000). The majority of metacognitive skills research has focused upon older children (age 9+) and adults, and therefore many of the methods used are heavily dependent on language and/or prior knowledge. This study aimed to more accurately characterize the metacognitive skills employed by 5 and 7-year-olds completing a problem-solving task, using observational coding of their verbalizations and non-verbal behaviors. This new approach allowed for two developmental issues in metacognitive skills to be addressed—first, whether developmental changes reflect quantitative or qualitative improvements (i.e. do older children do ‘more of the same’ as younger children, or employ different, more sophisticated skills?); second, whether metacognitive skills change with age, or with task-specific ability, or both. Here, we first present the theoretical model of metacognitive skills adopted in the current study. Next, we review current literature on the development of monitoring and control and highlight the related methodological issues.

### Theoretical model of metacognitive skills

There are a variety of models of metacognitive skill, and although not a strict dichotomy, these can be generally classed as either describing metacognitive components sequentially or describing the processes involved in metacognitive behavior regardless of timing. Sequential models of metacognition adopt a social cognitive perspective (as defined by Zimmerman 2001) and typically describe between three and five phases: ‘planning pre-task’, ‘metacognitive skills on-task’ and ‘evaluation post-task’ (Pintrich 2000; Winne and Hadwin 1998; Zimmerman 2000). In contrast, models that focus on the processes involved in metacognition adopt an information processing perspective on these behaviors and focus on feedback loops. Nelson and Narens’ (1990) influential model of metamemory processes<sup>1</sup> used during a task is one such model. They propose that cognitive processes are split into two or more specifically interrelated levels: the object-level is the actual cognitive task (e.g. completing a mathematical calculation), while the meta-level refers to a mental representation of the object-level, or cognitive task. Importantly, the direction of information flow between the levels represents either ‘monitoring’ (when information flows from the object-level to the meta-level) or ‘control’ (when information flows from the meta-level to the object-level). A simple example of monitoring is noticing a spelling error in a piece of writing; control processes would then allow the writer to use an alternative strategy to correct the error (e.g. attempting to ‘sound out’ the word). When coding the metacognitive skills used by young children, considering which direction the information is flowing proves useful for categorizing behaviors as monitoring or control.

### The development of monitoring and control

It is widely acknowledged that between the ages of 5 and 7 years a number of key developments in metacognitive skilfulness take place (Flavell et al. 1966, 1997; Keeney et al. 1967; Veenman et al. 2006), and some consideration has been given to the possible

<sup>1</sup> It should be noted that they did not attempt to incorporate declarative metacognitive knowledge in this model, and that it can also be broadened to encompass metacognitive skills used in a problem-solving task.

prerequisites and precursors of metacognitive skills. Likely candidates include theory of mind and executive functions. Of these, the relationship between metacognitive skills and executive functions has been more directly studied, but almost exclusively in adults (Bekci and Karakas 2006; Perrotin et al. 2007, 2008; Souchay and Isingrini 2004; Taconnat et al. 2009) making developmental conclusions hard to draw.

Aside from broad developmental trends there is surprisingly little experimental evidence regarding the early development of monitoring and control. The research findings generally converge to show that monitoring is fairly mature by age 6 (Butterfield et al. 1988; Schneider et al. 2000; Wellman 1977). However, Sperling and colleagues' (2000) finding that self-evaluation abilities reach maturity earlier than self-prediction abilities raises an important issue, as many studies utilize only one measure of monitoring which could mask some interesting developmental trends. Within the context of test-taking, it appears that monitoring skills are fairly established by age 9, whereas control skills (i.e. the ability to improve test scores by deleting incorrect answers) are still developing after age 9 (Roebbers et al. 2009).

All developmental research faces the challenge of disentangling age and ability effects, which requires task difficulty to be matched for different age groups. This issue is further complicated by the use of different definitions of ability; sometimes ability is conceptualized as a general ability or IQ, and other times ability is conceptualized as task-specific ability. Within metacognitive skills research, the challenge of disentangling age and ability appears to be more salient for control processes than for monitoring processes, as it has been found that the development of control may be accelerated in gifted children with superior IQs (results are mixed, but some indicate that the gifted advantage in control increases with age) whereas monitoring in gifted children only shows typical developmental improvements (Alexander et al. 1995). In addition, Eme and colleagues (2006) found no differences in monitoring accuracy between skilled and less skilled readers (an example of task-specific ability). Further, Puustinen (1998) found that in a sample of low and high achieving 7 and 10-year-olds (again, an example of task-specific ability), mathematics ability was more important than age in three aspects of self-regulation in a numerical problem-solving task. In the present study, no independent measure of general ability, or IQ, was collected. However, task-specific ability was manipulated by administering tasks of different difficulty levels. That is, the same task was more difficult for younger children, and less difficult for older children.

Some researchers have suggested that the interplay between monitoring and control processes drives the development of metacognitive skillfulness (Schneider et al. 2000). Before age 7, young children's poor performance on memory tasks (Schneider and Pressley 1997), and in a range of other domains (Hacker et al. 1998) is often considered to be caused by a breakdown in links between monitoring and control, where the information made available through monitoring processes is under-utilized in controlling subsequent performance (Schneider et al. 2000). Clearly, this relationship can only be understood if measures of both monitoring and control processes are collected in young children.

As elaborated upon below, there are some substantial methodological challenges in conducting metacognitive skills research in young children. Nevertheless, research has indicated that monitoring processes appear to reach maturity earlier than control processes, and control processes appear to be more related to both general and task-specific ability than are monitoring processes. However, there is limited research assessing naturally occurring metacognitive skills, which may provide a fairer measure of young children's abilities. This gap in the research literature could be filled by accurate online assessments of naturally occurring metacognitive skills, which encompass various aspects of the component processes.

## Methodological considerations

Many of the methods used in developmental metacognition research assume that participants have both mature communicative skills and conscious awareness of their own metacognitive skills, when in fact it has been shown that throughout development there is a fairly linear increase in the accuracy of self-reported strategy use (Winsler and Naglieri 2003) and that when using offline methods, which are collected prospectively or retrospectively, small changes to the methodology can have a considerable impact on the developmental conclusions drawn (Butterfield et al. 1988; Schneider et al. 2000). Indeed, there is little evidence available to suggest that a person's confidence judgements (CJs) or feelings of knowing (FoK) in a paired-associate learning task reflect the monitoring processes they use in everyday situations, particularly those that do not rely heavily on memory retrieval such as general problem-solving. In addition it is now widely accepted that online measures (those that are collected concurrently with task completion) more fairly represent the metacognitive skills that are truly being used by participants than do offline measures (Veenman et al. 2006), and that the think-aloud methodology is obstructive in children (van Hout-Wolters et al. 2000; van Someren et al. 1994).

Some metacognition researchers have begun to address the paucity of appropriate research methods for young children using naturalistic observational coding (Whitebread et al. 2005) or by conducting structured studies in order to make direct comparisons between age or ability groups (Molenaar et al. 2010). The latter approach, whereby participants are given the same problem-solving task to complete and their comments and/or non-verbal behaviors are coded for evidence of metacognitive skills, was taken in the current study. Notable absences of metacognitive skills were also coded. This makes an additional contribution to the literature, as the majority of research into metacognitive skills focuses on positive examples of metacognitive skills, and usually only one component process (either monitoring or control). However, video data collected here provided many examples of behaviors that indicated children had failed to use their metacognitive skills; these behaviors are reported in the literature as perseveration and distraction. Perseveration (Deak and Narasimham 2003; Morton and Munakata 2002; Zelazo et al. 2003) is considered to be a failure of control processes which appears behaviorally when a person persists with an incorrect response even when task rules or instructions can be recalled. Therefore, it reflects an inability to inhibit the original rule or behavior (rather than a memory or monitoring defect) in order to act flexibly and change behavior. Distraction is defined as an error of maintenance or monitoring (Barcelo and Knight 2002; Chevalier and Blaye 2008) as it stems from failing to maintain the task rules or one's position on-task (Chevalier and Blaye 2008). Interestingly, from 8 years of age until early adulthood errors of distraction are more prevalent than errors of perseveration but they show a different rate of development, with distraction errors reaching adult levels 2 years after perseveration errors (Crone et al. 2004).

## Current study

In summary, coding children's non-verbal behavior and verbalizations for evidence of metacognitive skills, while they complete a problem-solving task, offers a solution to many of the existing challenges facing researchers attempting to understand the development of metacognitive skills. These new methods may offer additional insights into two key issues in metacognitive skills development—do developmental changes in metacognitive skills reflect quantitative or qualitative improvements, and do metacognitive skills change with age or task-specific ability or both?

## Method

### Participants

Children were recruited from Year 1 (5-year-olds) and Year 3 (7-year-olds) classes at three schools in the city of Cambridge, UK. In total, 34 (19 female) children in the younger age group and 32 (10 female<sup>2</sup>) in the older age group agreed to participate, with equal proportions of multilingual children in each group (approximately one quarter). English was a primary language of all participants, we relied upon teacher report regarding language skills, and none of the participants had English as a second language. In one analysis, this sample size was reduced further to produce matched groups (elaborated upon in the relevant section).

### Tasks and variables





#### *The train track task*

Metacognitive skills were elicited using a quasi-naturalistic methodology—a controlled observation while children completed a problem-solving task. This task, which involves building a train track to match a predefined shape from a plan, was adapted from Karmiloff-Smith's (1979) closed-circuit railway task. Children were instructed to match a plan as best as they could, using as many pieces as required. This change from the original task (providing a plan of the shape to be constructed) allowed for a memory manipulation, where the children's memories were challenged when the plan was removed; it also allowed for certain aspects of monitoring to be evidenced more easily (e.g. checking the plan). Importantly, the task materials were familiar to the children (giving the task inherent appeal), but the task demands were novel.

The children attempted two shapes (one deemed 'easy' and one 'hard' for each age group, based on pilot data), and in one attempt the plan was removed to increase the memory demand of the task. In this study, 5-year-old children attempted an oval and a 'goggle' shape, and 7-year-old children attempted a 'goggle' and a 'P' shape (see Fig. 1). The order of shape, and the order of 'plan removal' were both counter-balanced to account for the possibility that children would show different behaviors on the second attempt as compared to the first, or for one particular shape, or in one memory condition. The task was introduced as follows: [In all conditions:] "In this game, we're not just playing with the train track, I'd like you to try and make some shapes. First of all I'd like you to try to make this shape [Plan presented] with the train track pieces." [Only in 'plan removed' condition:] "but I'm going to take this picture away, so make sure you have a good look at it first. Ok? [Plan removed when the child had looked for as long as they chose.] I'm going to take it away now." [In all conditions:] "So you can use as many pieces as you need, but you might not need all of them. You can spread out because it might be quite big, and I'd like you to tell me when you're finished." There was no experimenter interference in the task, if the child sought help only gentle encouragement was provided, and there was no time limit on the task. If they failed to state that they were finished at the end of the task, they were reminded "remember to tell me when you're finished". See Fig. 2 for the pieces of train track provided.

The videos of children attempting the train track task were subjected to analysis using two coding schemes—one to assess positive examples of metacognitive skills (i.e. behaviors that relate to higher scores on other measures of metacognitive skills and higher quality train tracks; Monitoring and Control), and one to identify failures of metacognitive skills

<sup>2</sup> Please note: Within each age group, there were no significant differences between the genders in any of the outcome scores.

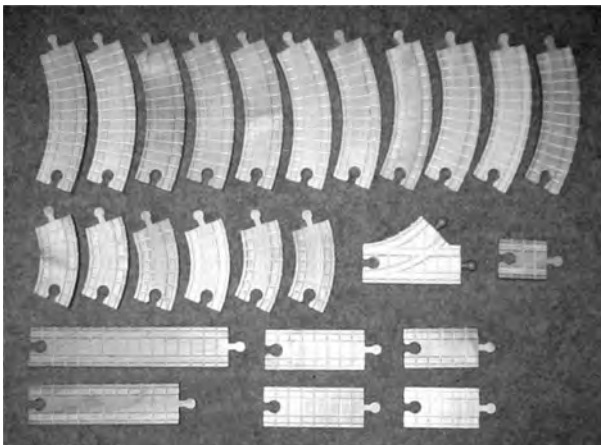
	Easy plans	Hard plans
Plans given to 5-year-olds	"Oval" 	"Goggles" 
Plans given to 7-year-olds	"Goggles" 	"P-shape" 

**Fig. 1** The train track plans attempted by 5-year-old and 7-year-old children

(Perseveration and Distraction). All video coding was conducted using Observer XT 9.0 software from Noldus. Further, the train tracks that children produced were scored for quality.

*Metacognitive skills coding scheme* The metacognitive skills coding scheme aimed to fairly represent, by numerical counts, positive examples of metacognitive skills demonstrated by children during the task. This was constrained by the decision to code only metacognitive skills in relation to cognitive activity, as opposed to motivation and emotional regulation. For instance, if a child showed evidence of trying to improve self-motivation during the task, this was not coded. Initially, this coding scheme was developed by identifying 21 verbalizations and behaviors from other coding schemes (Deloache and Brown 1987; Lambert 2001; Larkin 2000; Sangster 2010; Whitebread et al. 2009) and from a pilot study. As the present study adopted an information processing approach to metacognitive skills, the behaviors were categorized as either Monitoring (13) or Control (8) processes, adhering to Nelson and Narens' (1990) model. A Monitoring behavior is defined as one that serves to update the mental representation of the task, while a Control behavior asserts some action at the level of the task (see Table 1 for code descriptions, examples, and rates).

To aid understanding of the metacognitive skills coding scheme, and to explicate how we have applied the Nelson and Narens' (1990) model to the wide array of metacognitive behaviour observed in young children, two codes, Reviewing and Planning, are further elaborated upon and their categorization as Monitoring and Control processes, respectively,



**Fig. 2** Train track pieces available to the children

**Table 1** The metacognitive skills coding scheme. Codes that were later dropped from analysis are identified by an asterisk (\*). Total (across both tasks) mean (standard deviation) rates per minute are also provided

Code name and description	Example	Mean (SD) rates
<b>Monitoring behaviors</b>		
Checking Own, a pause to review whole of own construction. Not checking only one area.	Child pauses and looks at whole train track they have made so far.	2.02 (1.12)
Checking Plan, checking the original train track plan.	Child glances back to plan of train track they are working on.	1.90 (1.80)
Prospective Monitoring, judgement of task before task begins.	<i>This is going to be a challenge!</i>	0.09 (0.21)
Clarification, clarifying task demands. Can be before or during task.	<i>Do I use all the pieces?</i>	0.05 (0.14)
Reviewing, a pause to look at different pieces before or during task, not seeking any particular piece. Lasts for at least 3 s.	Child glances around all the pieces of train track.	0.44 (0.51)
Self-questioning, child highlights a problem to be solved, poses themselves a question.	<i>How will it curve around?</i> (to self)	0.45 (0.15)
*Task Difficulty, child comments on task difficulty during task.	<i>This one is actually very difficult!</i>	0.14 (0.31)
*Error Detection, a pause or comment that clearly shows they have noticed an error. Could follow checking or be spontaneous.	Child pauses and makes a face while looking at problem area. <i>That's not right!</i>	0.62 (0.69)
Commentary, comment on success, what has been achieved so far.	<i>There, that's a tiny bit better. Right, that bit is done...</i>	0.39 (0.85)
*Using Other for Monitoring, identifying help is needed but not asking, looking to experimenter.	<i>I can't do this bit</i> (child looks to experimenter but does not ask for help).	0.27 (0.48)
Evaluation, spontaneous evaluation of product at the end of the task.	<i>But that bit isn't right.</i>	0.04 (0.11)
Justified Termination, child announces task is finished, without being prompted, is correct.	<i>There. Finished!</i>	0.43 (0.47)
Memory Monitoring, child comments or indicates that they are struggling with their memory.	<i>Oh, I can't remember...</i> Makes a 'memory' face.	0.06 (0.19)
<b>Control behaviors</b>		
Clearing Space, child clears space in which to work or starts working in an already empty space.	Child clears a space on the table with hands before placing first piece down.	0.54 (0.56)
Planning, explicitly stating a plan. Can be before, or during the task.	<i>I'm going to do these straight bits first.</i>	0.54 (0.80)
Sorting, organising or grouping materials before and during task. Includes adding pieces symmetrically.	Child compares the lengths of two straights.	0.46 (0.76)
Seeking, searching for materials before and during task. Lasts for at least 2 s.	Child searches for large curve. <i>Do you have straights?</i>	1.13 (0.94)
*Requesting Help, realising at an appropriate point that they cannot do something and asking for help.	<i>Can you help me join them up?</i>	0.02 (0.08)
Change Strategy, using a different strategy or piece than before. Not just first strategy or piece chosen.	Child reverses piece of track so that it curves in the correct way.	2.04 (1.06)
*Gesture, the use of gesture to support cognitive activity or communication.	Child traces the shape they need with their finger.	0.33 (0.53)
Memory Aid, the use of a memory aid or labeling the train track shape.	<i>It's the shape of B</i> (about goggle shape).	0.08 (0.22)

are justified. Reviewing is a monitoring behavior because it involves a person updating their representation of the object-level. Information ('what pieces are available') flows from the object-level (the task) to the meta-level (the mental representation of the task). Importantly, this behavior asserts no action at the object-level, as is the definition of a control behavior. Planning as defined here (the child explicitly stating their intentions) is a control behavior because this reflects information flowing from the meta-level to the object-level. Explicitly stated planning is the manifestation of the meta-level asserting some action on the object-level. The child has considered the task, monitored where they are in the task, and now they are stating their next move or approach to the task—this is the action at the object-level.

As opposed to interval coding, the behaviors were coded whenever they occurred (as the behaviors or comments had to be viewed in context and some codes were applied to a sequence of behaviors, rather than a distinct unitary behavior). The codes were then transformed into rates, by dividing the number of occurrences by the number of minutes the child spent on the task. There were two levels of scores produced using this metacognitive skills coding scheme: macro-level behavioral coding rates (aggregate rates of all individual behaviors that contribute to Monitoring or Control processes), and individual behavior code rates. As well as considering the rates of individual behavior codes, the coding data were converted into categorical data by considering whether each participant showed the behavior at all during the attempt to build the shape. These were then tested using a chi-square test for differences between the age groups, and provided slightly different information regarding the children's metacognitive abilities. The first stage of data processing sought to identify which individual behavior codes should be retained as 'positive examples of metacognitive skill' and this resulted in five of the individual behavior codes being dropped from further analysis (identified by asterisks in Table 1).

*Perseveration and distraction coding scheme* Examples of Perseveration and Distraction were generated from the pilot study data (see Table 2 for code descriptions and examples). Instances of Perseverative behaviors and Distraction were summed and transformed into rates, to produce macro-level Perseveration and Distraction rates. The mean (and standard deviation) of Perseveration and Distraction total rates (across both tasks) were 1.14 (1.39) and 0.10 (0.25), respectively.

Two levels of inter-rater reliability for the two coding schemes combined ('metacognitive skills' and 'perseveration and distraction' coding schemes) were calculated (Bakeman and Gottman 1997)—agreement that a timepoint should be coded (named Unitizing agreement), and agreement on how that timepoint should be coded (named Coding agreement). The Cohen's Kappa ( $\kappa$ ) statistic can only be calculated on the Coding agreement. The inter-rater Unitizing agreement was 59 %, inter-rater Coding agreement was 90 % and Coding agreement  $\kappa=0.90$  (these levels of agreement are considered acceptable for this type of data, according to Bakeman and Gottman 1997). Further, intra-rater reliability was calculated by the primary researcher coding 10 % of the videos twice with a minimum gap of 2 weeks between coding occasions. The intra-rater Unitizing agreement was 85 %, intra-rater Coding agreement was 98 % and Coding agreement  $\kappa=0.98$ .

*Train track quality score* The quality of the children's finished train tracks were scored according to how well they matched the shape in the plan, by identifying important features of the tracks. It was also possible to have points deducted for clear errors (e.g. obvious extra pieces). Coincidentally, each train track had a maximum quality score of six. Full details of how quality was scored are available in Online Resource 1. A second coder scored the

**Table 2** The perseverance and distraction coding scheme

Code name and description	Example
<b>Perseveration behaviors</b>	
No Strategy, when something will not work, uses same strategy over and over or gives up. NB: one code given for repeated use of the same strategy.	Child tries to force two pieces of train track over and over and then gives up.
Not Following Plan, making a plan and then not following it.	Child says they will start with straight edge, but then places a curved track piece.
Narrow View, a failure to review the whole area or whole array of train track pieces.	Child looks for a straight piece but only looks in one area of the table.
Focus on Join, focusing on joining up the circuit rather than making the shape.	Child is making a circle, but is so focused on making it join that s/he adds straight pieces.
Two Positives, when child ends up with two sticking out ends, not realising they need to reverse one.	Child needs to add one more 'junction-piece' (it is in the wrong position and just needs to be reversed). Child does not succeed.
Large/Small Curves, when they run out of small curves, not realising they need to switch to large curves.	In making a curved edge, child uses all small curves and has not realized s/he needs to switch to large curves.
Finishing Error, saying they are finished when there is a major discrepancy between their track and the plan.	Child says s/he has finished, when s/he has made a circle, not the goggle-shape. Meant to be a straight but has curves.
Goal Neglect, showing awareness of the rule/error but not acting accordingly.	<i>It's 'cause these ones don't turn as well as those ones [but does not fix it].</i>
<b>Distraction behaviors</b>	
Off-task, child seems to have gone completely off task and forgotten goal.	Child seems to be just playing with the train track and not attempting to match the plan.
Distracted, child is distracted by other people's activity or own thoughts, but works on the task intermittently.	Child keeps looking at a class outside. <i>"My aunt's picking me up today".</i>

quality of 130 train tracks and achieved good inter-rater agreement (goggle shape  $\kappa=0.85$ , oval shape  $\kappa=0.88$ , p-shape  $\kappa=0.88$ ).

In conclusion, the train track task elicited a range of variables. First, rates of positive metacognitively skilled behaviors (rates of macro-level Monitoring and Control, and individual behavior codes); second, rates of a lack of metacognitive skills (Perseveration and Distraction rates); third, the quality of the train track. As each child completed two train tracks, each of these measures was available for 'easy' and 'hard' plans, and for when plans were 'removed' or 'available'. A total score was also produced by summing the children's rates or scores across both plans.

### *The CHILD questionnaire*

The Children's Independent Learning Development (CHILD) questionnaire (Whitebread et al. 2009) was collected from class teachers about each child, to obtain an alternative view of their metacognitive skills and to aid validation of the metacognitive skills coding scheme. Teachers were asked to indicate whether children never, sometimes, usually, or always show certain behaviors. A reliability study of the CHILD questionnaire using exploratory factor analysis, with a sample of 423 children aged 4 to 6 years, produced a two factor structure

(Whitebread et al. 2011). The self-regulation factor (9 items) represents broad, cognitive aspects of metacognitive skills; the social-regulation factor (4 items) concerns more pro-social aspects of the checklist. The internal consistency of each factor, as measured by Cronbach's alpha within the present study sample, was high: self-regulation factor  $\alpha=0.96$ , social-regulation factor  $\alpha=0.86$ . In summary, there are two variables from this measure—the mean self-regulation score, and the mean social-regulation score.

### Normality of variables and approach to analysis

Once data from the train track task were prepared, the suitability of the data for parametric statistical analysis was assessed, both as a whole group and in each age group separately. Normality was assessed by the Kolmogorov-Smirnov test of deviation from the normal distribution; homogeneity of variance across the two age groups was assessed by Levene's test. As the distribution of many variables violated the assumptions required for parametric tests, a number of transformations were applied to the data: the natural log, log<sub>10</sub>, reciprocal and square root (as reported in Field 2009, p. 153–164). Of these, the square root transformation had the largest impact and is reported here, although some variables remained not normally distributed (train track quality score, Perseveration rate and Distraction rate). As the measures of 'failures' of monitoring (Distraction) and control (Perseveration) were not normally distributed, their composite (named 'PD') was used in all further analysis as it was normally distributed. Further, in all analyses, both parametric (on square root transformed data) and nonparametric (on raw data) tests were used where possible. In correlations, age was partialled in the parametric analyses. To ameliorate the risks associated with transformed data, only results that were at least marginally significant using both parametric and nonparametric methods were accepted, and nonparametric values are reported in the text. This procedure is acknowledged to reduce the risk of type I errors (Field 2009, p. 551). Effect sizes are reported (Cohen's *d*) where appropriate, and were calculated using software created by Devilly (2005).

## Results

### Validating the metacognitive skills coding scheme

The metacognitive skills coding scheme was intended to measure positive indicators of metacognitive skills, meaning that higher rates of these behaviours would relate to higher scores on other metacognitive measures and higher quality end products. Of course, sometimes the same behaviour can be positive in one circumstance and debilitating in another circumstance. For this reason, the context was always considered when applying these coding schemes, and an appropriate code from the Perseveration and Distraction coding scheme was applied when a child's behaviour became inflexible and/or debilitating. To clarify which individual behavior codes from the metacognitive skills coding scheme were positive indicators of metacognitive skills, and therefore should be included in the macro-level Monitoring and Control rates, two types of analyses on the individual behavior codes were conducted: first, correlations among scores on the CHILD questionnaire and individual behavior code rates; second, correlations among the train track quality scores and individual behavior code rates. Both total rates (shown in easy and hard plans combined) and rates in easy and hard plans separately were considered. Only significant correlations ( $p<.05$ ) are presented in the text.

*Associations among CHILD questionnaire and individual behavior codes* As the metacognitive skills coding scheme was designed to be a positive approach to coding (i.e. higher rates of any individual behavior would indicate better metacognitive skills) the association between individual behavior codes and the two factors of the CHILD questionnaire were investigated. Overall, five individual behavior codes were found to be positive indicators of metacognitive skills as they increased with scores on the CHILD. These were: Checking Own (with self-regulation factor, correlation coefficients ranged from  $r=.32$  to  $.46$ ), Checking Plan (with social-regulation factor  $r=.38$ ), Reviewing (with social-regulation factor  $r=.45$ ), Error Detection (with both factors  $r=.32$  to  $.56$ ), and Memory Monitoring (with social-regulation factor  $r=.30$  to  $.41$ ). The only individual behavior code that showed negative correlations with either factor of the CHILD questionnaire was Gesture (with both factors  $r=-.37$  to  $-.55$ ). This suggests that Gesture is not a positive indicator of metacognitive skills.

*Associations among train track quality and individual behavior codes* Clearly, with the understanding that metacognitive skills are critical for educational or task success, it becomes important to understand which specific behaviors are related to producing (in this case) a high quality train track. The individual behavior code Justified Termination had the most consistent and highest correlations with train track quality (correlation coefficients ranging from  $r=.45$  to  $.71$ ), which is unsurprising as a high quality train track was almost a prerequisite of receiving this code. However, all the other individual behavior codes showing significant correlations with train track quality were process oriented, and could be understood to make a direct contribution to train track quality.

Four other Monitoring and four Control codes were associated with higher quality train tracks. The Monitoring codes were Checking Plan (with hard plan in older children,  $r=.59$ ), Prospective Monitoring (with hard plan in all children,  $r=.38$ ), Clarification (with hard plan in all children,  $r=.25$ ), and Commentary (with hard plan in all children,  $r=.29$ ). The Control codes were Clearing Space (with easy plan in all children,  $r=.44$ ), Planning (with easy plan in younger children,  $r=.35$ ), Sorting (with hard plan in all children,  $r=.30$ ) and Seeking (with hard plan in older children,  $r=.33$ ). These individual behavior codes could result in higher quality train tracks for a range of reasons: some indicated the child was aware of the challenge that the task posed and responded appropriately (e.g. Checking Plan, Prospective Monitoring, Clarification), some reflected heightened self-monitoring during the task (e.g. Commentary, Justified Termination), some suggested that the child was anticipating the train track shape and taking the appropriate preparatory steps (e.g. Clearing Space, and Planning in the younger group) and some indicated heightened control processes during the task (e.g. Sorting, Seeking).

Some negative correlations indicated that certain behaviors were associated with poorer quality train tracks. These were Task Difficulty (with easy plan in all children,  $r=-.24$ ), Error Detection (with easy plan in older children,  $r=-.65$ ), Using Other for Monitoring (with easy plan in all children,  $r=-.44$ ), Planning (with easy plan in older children,  $r=-.39$ ) and Gesture (with hard plan in older children,  $r=-.30$ ). The three Monitoring codes (Task Difficulty, Error Detection, Using Other for Monitoring) all indicated that the children were experiencing difficulty, and had sufficient metacognitive skills to be aware of the difficulty or errors. Therefore, the experience of difficulty or challenge is most likely the mediating factor in these correlations. The negative correlations involving Control codes (Planning and Gesture) are harder to interpret. However, the fact that these correlations were only significant in older children could hold some explanatory power—that is, perhaps older children who needed to use gesture in the hard plan or planning in the easy task were experiencing difficulty. In turn, the fact that they were finding the task challenging may be what causes them to produce a poor quality train track, and these Control codes may merely reflect task

difficulty. This interpretation is supported by the finding that a higher percentage of younger children than older children demonstrated both of these individual behaviors, so these behaviors perhaps indicate immaturity in the older group.

This analysis also revealed some developmental differences between the age groups regarding which behaviors relate to high quality train tracks. For instance, Task Difficulty was positively associated with train track quality in the younger group attempting the hard plan ( $r=.32$ ), but negatively associated with train track quality in the older group attempting the easy plan ( $r=-.32$ ). Clearly, this indicates that the hard plan for the younger group was indeed an appropriate challenge, and those children who recognized this were able to produce higher quality tracks. Conversely, the easy plan was possibly too easy for the older children and therefore those children who commented on it being difficult may be less skilled at making train tracks and ultimately produce poorer quality tracks. There was a similar finding with regard to Planning in the easy task—in the younger children this was associated with higher quality tracks ( $r=.35$ ), probably because this train track was an appropriate challenge for them, whereas Planning during the easy task for the older group was associated with poorer quality train tracks ( $r=-.39$ ), as interpreted above.

*Summary: validating the metacognitive skills coding scheme* In conclusion, the foregoing analysis identified five codes that were consistently related to poor metacognitive skills or quality scores, and therefore should not be included when composing the macro-level Monitoring and Control scores. Three of these were from the Monitoring group—Task Difficulty, Error Detection and Using Other for Monitoring—and two were from the Control group—Requesting Help and Gesture. The three Monitoring codes could be interpreted as reflecting ‘helplessness’. That is, they involve commenting on failure and seeking other-regulation, as opposed to being purposeful and self-regulating. Interestingly, when these three codes were combined to produce a composite score, they seemed to be appropriate behaviors for the younger group (as they showed positive correlations with the CHILD scores), but inappropriate for the older group (as they showed negative correlations with the CHILD scores and train track quality). However, since the coding scheme was intended to reflect positive metacognitive skills across the 5 to 7-year-old age range, these were removed. Among the Control codes, the behavior code Requesting Help was removed as it was deemed to also reflect a certain level of helplessness and showed different patterns across the age groups. Gesture was removed as it consistently showed relationships with poor quality train tracks and low scores on the CHILD.

Following this analysis and removal of codes, there remained ten individual behavior codes that contributed to the macro-level code Monitoring (Checking Own, Checking Plan, Prospective Monitoring, Clarification, Reviewing, Self-questioning, Commentary, Evaluation, Justified Termination, Memory Monitoring) and six that contributed to the macro-level code Control (Clearing Space, Planning, Sorting, Seeking, Change Strategy, Memory Aid). Of those remaining in each category, individual behavior codes that were shown by fewer than 25 % of children in each group (Prospective Monitoring, Clarification, Self-questioning, Evaluation, Memory Monitoring, and Memory Aid) were still included in the macro-level codes but were not analyzed in terms of individual behaviors due to the limited range of scores. There was not enough variance in these scores to produce statistically meaningful results, as 75 % of the children had rates of zero in these individual behavior codes.

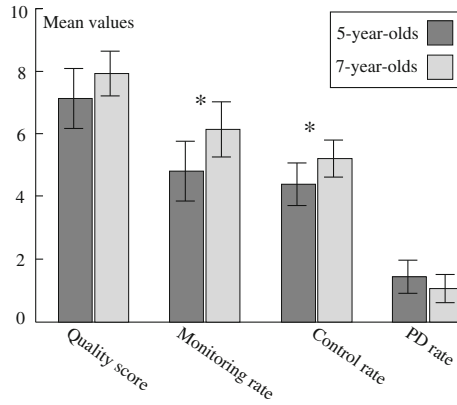
## The development of metacognitive skills

Two research questions were addressed using this new methodology. First, were the developmental changes shown in metacognitive skills between 5 and 7-year-olds quantitative and/or qualitative? Evidence for this came from comparing the behaviors demonstrated by each age group when task difficulty was matched (total rates across both plans). The second question was, do metacognitive skills improve with age or task-specific ability or both? Evidence for this was provided from the rates of behaviors shown when task difficulty was not matched between the age groups (rates on the goggle plan). To clarify, comparing the children on the same plan (the goggle plan) is one way of examining the effects of task difficulty or task-specific ability, as this shape was ‘hard’ for the younger group and ‘easy’ for the older group. In this respect, task difficulty can be seen as the inverse of task-specific ability, as if one experiences a task as being more difficult then, by definition, they have lower task-specific ability. Comparing the children on behaviors shown across both plans was intended to eliminate the effects of task difficulty or task-specific ability, as each group attempted one ‘easy’ and one ‘hard’ plan. Since the teacher reported metacognitive skills measure, the CHILD questionnaire, was based on how a child compared to their peers (rather than an objective level), no differences in scores were expected between the two age groups. Indeed, neither factor showed a significant difference. However, this finding is important as it indicates that the groups were normal for their age in metacognitive skillfulness—i.e. the younger group did not have, by chance, unusually high metacognitive skills, in comparison to their peers.

*Do developmental changes in metacognitive skills between 5 and 7-year-olds reflect quantitative or qualitative improvements?* As seen in Fig. 3, the groups produced train tracks of a similar quality (i.e. no significant difference between the age groups on train track quality), indicating that a similar level of difficulty was experienced by each group. Within the macro-level codes, the rate of Monitoring was significantly different between the age groups ( $U=360.0, p=.02$ ; effect size  $d=0.51$ ) as the older group showed higher rates (6.14/min) than the younger group (4.83/min). Control rates also showed a significant difference between the age groups ( $U=389.0, p=.047$ ;  $d=0.45$ ), whereby the older group showed higher rates of Control (5.21/min) than the younger group (4.41/min). There was no significant difference on the rate of PD.

One individual behavior code that contributes to the Monitoring rate showed a significant difference between the age groups—Checking Plan ( $U=334, p=.01$ ;  $d=0.49$ ), which older children demonstrated over 50 % more frequently (2.34/min) than younger children (1.48/min). Interestingly, the types of checking appear to show a dissociation across ages—that is, the older children checked the plan more often than they checked their own construction (Checking Plan: 2.34, Checking Own: 2.16/min), while the younger children checked their own construction more often than they checked the plan (Checking Plan: 1.48, Checking Own: 1.88/min). Although both are positive metacognitive skills, checking their own construction may be a less mature skill than checking the plan. The older group also showed significantly higher rates of two individual behavior codes that contribute to Control: Clearing Space ( $U=259, p<.001$ ;  $d=0.96$ ; 5-year-olds: 0.30 vs. 7-year-olds: 0.79/min) and Seeking ( $U=328, p=.006$ ;  $d=0.61$ ; 5-year-olds: 0.87 vs. 7-year-olds: 1.42/min). See Fig. 4 for bar charts representing these findings.

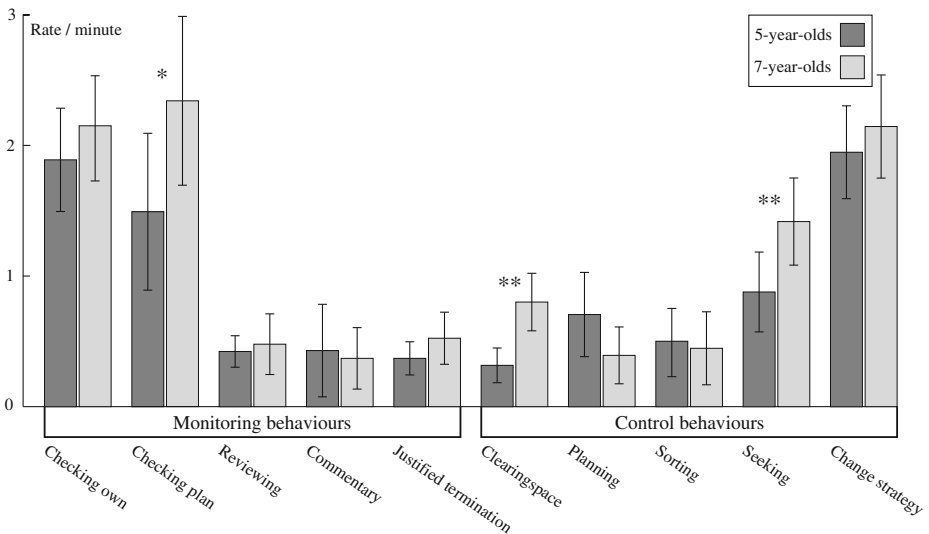
When the categorical data (regarding whether a participant showed or did not show the behavior) across both tasks were analyzed, four individual behaviors showed significant associations between age group and whether the behavior was displayed. These were



**Fig. 3** Mean train track quality scores and rates of macro-level metacognitive skill codes shown across both plans (total rates per minute). Error bars represent 95 % confidence intervals. Significant difference between the age groups: \* $p < .05$

Justified Termination ( $\chi^2(1)=8.05, p=.01$ , showed by more older children), Clearing Space ( $\chi^2(1)=9.83, p=.002$ , showed by more older children), Planning ( $\chi^2(1)=3.76, p=.05$ , showed by more younger children), and Seeking ( $\chi^2(1)=3.13, p=.08$ , showed by more older children).

In summary, examining the macro-level codes of Monitoring and Control showed quantitative increases in metacognitive skills in this age range—7-year-olds showed higher rates of both Monitoring and Control than 5-year-olds. However, examining the individual behavior codes also showed qualitative changes in this age range. That is, the individual Control behavior codes that showed differences between the groups reflected planning behaviors—Clearing Space and Seeking were more common in the older group, whereas



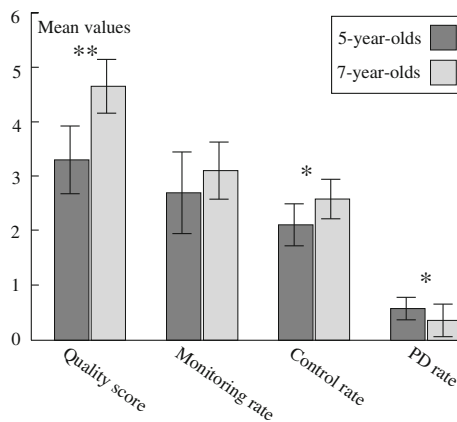
**Fig. 4** Mean rates of individual behaviour codes shown across both plans (total rates per minute). Error bars represent 95 % confidence intervals. Significant difference between the age groups: \* $p < .05$ , \*\* $p < .01$

verbally expressed Planning was displayed by more of the younger than the older children. Further, although these results did not reach significance, this data could indicate an interesting trend regarding the development of monitoring—where younger children check their own construction more and older children check the plan more. Thus, this approach to assessing metacognitive skills in young children was sufficiently developmentally sensitive to identify both quantitative and qualitative changes between 5 and 7 years of age.

*Do metacognitive skills change with age or task-specific ability or both?* In this analysis, children from each age group were compared on the same plan (the goggle shape). Therefore, it was necessary to match the number of plans in each memory condition (those in which the plan was removed or available). For this purpose, five children in the ‘plan removed’ condition were eliminated from the 5-year-old group and three children in the ‘plan available’ condition were eliminated from the 7-year-old group. They were randomly selected using a random number generator. After this process, in each age group there were 14 videos where the plan was removed and 15 where the plan was available. There were no significant differences between the data eliminated and those which remained on any coding rate; this indicates that the random selection of videos to be eliminated did not distort the mean of either age group. Analysis was completed on two levels—the macro-level and individual codes.

The older children produced significantly higher quality goggle train tracks (4.76) than the younger children (3.31;  $U=228.5$ ,  $p=.002$ ;  $d=0.91$ ), confirming that this comparison informs us about behaviors shown by the two age groups when difficulty is not matched. Two of the macro-level metacognitive skills codes also showed significant differences between the age groups. The rate of Control ( $U=287.0$ ,  $p=.04$ ;  $d=0.48$ ) was higher in the older group (2.77/min) than the younger group (2.22/min). PD rates were also significantly different between the age groups ( $U=279.0$ ,  $p=.02$ ;  $d=0.27$ ) whereby the younger children showed higher rates (0.58/min) than the older children (0.39/min). See Fig. 5 for a graphical comparison of train track quality scores and rates of macro-level metacognitive skill codes.

The rates of individual codes were also examined to determine which behaviors differed within the broader categories of monitoring and control. In Monitoring codes,

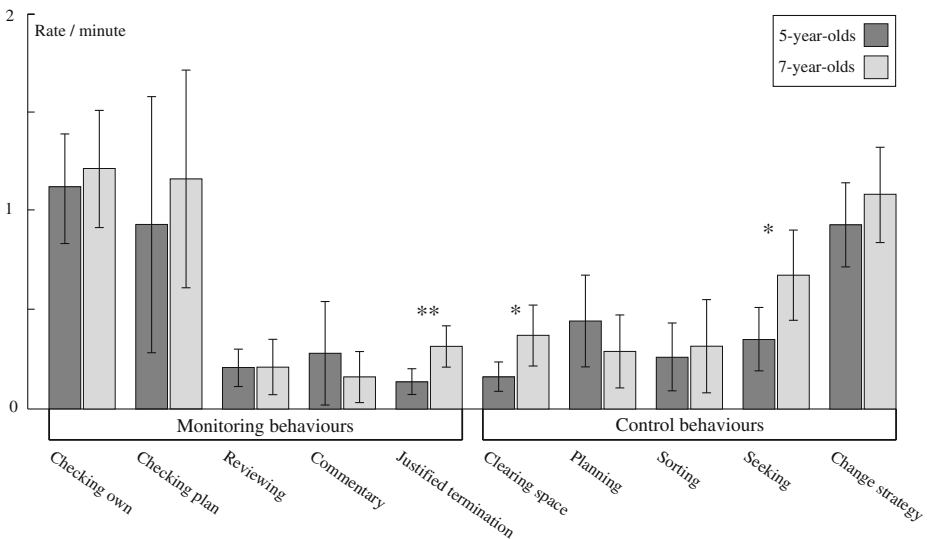


**Fig. 5** Mean train track quality scores and rates of macro-level metacognitive skill codes shown during the goggle plan (rates per minute). Error bars represent 95 % confidence intervals. Significant difference between the age groups: \* $p < .05$ , \*\* $p < .01$

only one behavior showed a significant difference, which was the ability to make a Justified Termination of the task ( $U=272$ ,  $p=.02$ ;  $d=0.74$ ); older children demonstrated this behavior more frequently (0.30/min) than younger children (0.12/min). In Control behaviors, Clearing Space ( $U=302$ ,  $p=.06$ ;  $d=0.63$ ) occurred significantly more frequently in the older group (0.37/min) than the younger group (0.16/min). Likewise, there was a significant difference between the age groups on the rate of Seeking behavior ( $U=289$ ,  $p=.003$ ;  $d=0.48$ ), where older children showed more seeking behaviors (0.64/min) than younger children (0.37/min). Figure 6 shows these results.

The chi-square test on categorical data (indicating whether each participant showed the behavior at all during the task) resulted in the same individual behavior codes showing significant differences between the age groups: within Monitoring, more older children showed Justified Termination ( $\chi^2(1)=5.20$ ,  $p=.02$ ); within Control, more older children showed Clearing Space ( $\chi^2(1)=4.06$ ,  $p=.04$ ) and Seeking ( $\chi^2(1)=4.61$ ,  $p=.03$ ). In addition, more younger children showed PD ( $\chi^2(1)=6.01$ ,  $p=.01$ ).

This analysis demonstrated that when children were attempting the same problem-solving task, which was more difficult for younger than older children (as shown by the train track quality scores), the rate of Control and the presence of negative behaviors (PD) emerged as showing differences between the age groups. This confirms the value of coding failures of metacognitive skills, and not only positive behaviors. Further, within Monitoring behaviors only the ability to terminate the task appropriately showed differences between the age groups (which was possibly driven by differences in train track quality), while two Control codes (Clearing Space and Seeking) showed significant differences between the age groups.



**Fig. 6** Mean rates of individual behaviour codes shown during the goggle plan (rates per minute). Error bars represent 95 % confidence intervals. Significant difference between the age groups: \* $p < .05$ , \*\* $p < .01$

## Discussion

The research presented here addressed two major issues regarding developmental change in metacognitive skills—first, are developmental changes purely quantitative (i.e. older children do ‘more of the same’ as younger children) or are there qualitative improvements in metacognitive skills? Second, do metacognitive skills change with age, or with task-specific ability, or both? The findings will be summarized in order, beginning with how monitoring develops, followed by how control develops, how PD changes, and concluding with the impact of task-specific ability on metacognitive skills.

The majority of previous literature on the development of metacognitive skills indicates that monitoring processes (as reflected in FoK or CJ accuracy) are fairly mature by 6 years of age, and certainly earlier to mature than control processes (Butterfield et al. 1988; Krebs and Roebers 2010; Roderer and Roebers 2010; Roebers et al. 2009; Schneider et al. 2000; Sperling et al. 2000). The present study showed that when task difficulty was matched, the quantity of monitoring behaviors increased between 5 and 7-year-olds. Further, due to the in-depth coding scheme used in this study, qualitative changes also emerged. Within Monitoring codes, the older children checked the train track plan over 50 % more often than the younger children, while younger children checked their own constructions more often than they checked the plan. Correlations with train track quality showed that checking the plan was only related to high quality train tracks in the older group. These data can be interpreted in two ways—either, they could suggest that if the younger children checked the plan more often they would produce higher quality tracks (and a lack of correlation in the younger group is due to a lack of variance in the rates), or that 5-year-olds would not benefit from checking the plan more often due to an immaturity in the interplay between monitoring and control processes (an issue highlighted by Schneider et al. 2000). The other Monitoring behavior that increased with age when task difficulty was matched was Justified Termination—reflecting a child’s ability to evaluate their task success and appropriately choose to end the task. It is interesting that the two monitoring behaviors that seem to reflect developmental maturity in this task are quite different—that is, Checking Plan occurs on-task and could be considered a ‘basic’ monitoring skill, whereas Justified Termination occurs at the end of the task and serves a more complex evaluative purpose. In sequential models of metacognition these behaviors are considered to contribute to different phases. Further, making a justified termination of the task may partly reflect the ‘interplay’ between monitoring and control (according to Schneider et al. 2000). That is, justified termination requires the participant to both monitor and act appropriately on the information received. That both a basic monitoring skill and a more complex one increase in frequency in this age range suggests a general development across different types of monitoring behaviors.

The older group also demonstrated more metacognitive control than 5-year-olds when task difficulty was matched. Again, when individual behaviors were examined, qualitative changes also appeared regarding the type of planning that takes place; these changes may reflect the decline within this age group in the use of what is now commonly referred to as ‘private’ or ‘self-directed’ speech (Winsler et al. 2009). That is, young children’s planning was explicitly stated (receiving the Planning code), whereas older children’s planning was reflected in more internalized preparatory behaviors (receiving Clearing Space and Seeking codes). We found that, in turn, these specific behaviors contribute to higher quality end products—Planning is positively associated with high quality train tracks in the younger group, but is negatively associated with train track quality in the older group, while Clearing Space is related to high quality train tracks in children of each age, and Seeking is related to high quality train tracks only in older children. The individual behavior codes that change

with age reflect on-task behavior that increasingly shows a clear internal plan being followed and a goal being maintained throughout the task, as well as increasingly superior evaluative abilities. The present study indicates that both monitoring and control processes increase quantitatively and improve qualitatively in this age range. These findings are distinct from previous developmental studies of metacognitive skills, which have tended to examine specific aspects of monitoring or control processes, in memory tasks.

While both monitoring and control behaviors showed developmental increases in this age range when task difficulty was matched, the rate of PD (the ‘lack’ of metacognitive skills) did not decrease with age. However, the rate of PD was affected by task difficulty (which can also be interpreted as ‘task-specific ability’). When completing the same task, more 5-year-old children showed PD behaviors than 7-year-olds, and younger children also showed higher mean rates of PD than older children. These results illustrate the value of coding negative indicators of metacognitive skills as well as positive ones, as these behaviors appear to be more related to task-specific ability or task difficulty than to age. Task-specific ability also appeared to affect control processes—as when children were attempting the same task, older children (with better abilities) demonstrated significantly higher rates of Control than younger children (with poorer abilities). This supports evidence from the literature that control processes are more affected by both general and task-specific ability than are monitoring processes (Alexander et al. 1995; Eme et al. 2006; Puustinen 1998). To be more precise, the present study indicated that rates of Monitoring behaviors increase with age, Control behaviors increase in frequency with both age and task-specific ability, and PD behaviors decrease in frequency with increasing task-specific ability.

The approach taken here to describe young children’s metacognitive skills allowed for an understanding of how both the quantity and quality of metacognitive behaviors change with age, and improved upon offline methods such as CJs. The present study aimed to address each of the challenges faced by prior research in this area. First, children were not asked to give verbal reports while participating in the train track task. If they did speak, their verbalizations were coded, but it was not a requirement and the majority of the codes given were for non-verbal behavior. This minimized the bias towards older and more verbally able children. Second, in this study the focus was not on memory processes, but instead on the metacognitive skills used in a fairly naturalistic problem-solving task. The problem-solving train track task was developed in order to more accurately represent the metacognitive skills that young children use in an environment that is meaningful to their everyday lives, as they were set a novel challenge with familiar materials. The challenge itself was intended to make limited demands on prior knowledge, and was inherently appealing to the children. Third, in the present study, an attempt was made to match task difficulty across age groups by selecting train track plans based on how children of the same age performed on these tracks previously. Each group attempted one track that was deemed ‘easy’ and one that was deemed ‘hard’ for their age group. This allowed for a systematic comparison of the effects of task difficulty (and therefore task-specific ability). Finally, both monitoring and control behaviors were assessed in the present study, as well as behaviors that indicated a ‘lack’ of metacognitive skills (rates of PD).

### Limitations

It must be recognized that the metacognitive skills coding applied to the train track task was fairly constrained; that is, the focus was on metacognitive skills as they applied to a cognitive task. Any evidence of a child monitoring and controlling their emotions or motivation during the task was not analyzed. Instead, an attempt was made to make the train track task inherently motivating so that motivation would not cause the children a

significant challenge. Clearly, the ability to monitor and control one's emotions and motivation are important determinants of how appropriately children use their metacognitive skills, which were not accounted for by the present metacognitive skills coding scheme.

While there are obvious advantages to reducing the verbal requirements of a metacognitive skills assessment, one drawback to the approach taken in the present study is that there was no opportunity to ask the children to elaborate on their behavior or comments. In particular, this may have affected our ability to observe evaluation skills. An adaptation that may improve upon this limitation is to ask carefully structured supplementary evaluative questions at the end of the task. Although this would of course be an offline method and the responses would be prompted as opposed to naturally occurring, it would be interesting to examine whether these online and offline measures of monitoring corroborate one another. Another issue to be considered is whether these developmental findings are specific to this particular task, or if they can be generalized. It is not possible to determine this from the present study, and therefore the findings presented here can only be understood within the framework of the train track problem-solving task. However, it would be simple and valuable to apply this coding scheme to another task and examine the metacognitive skills shown across two tasks. Any other task should have two key features in common with this one—it should make minimal demands on prior knowledge or verbalizations, and a plan should be provided (because in a practical sense it is useful to have participants check something physical rather than a mental representation).

The risks of drawing developmental conclusions solely from cross-sectional designs are well known and reviewed elsewhere (Schneider et al. 2009). An additional outcome of using a cross sectional design is that the sample size is reduced (i.e. two groups of approximately 30 participants in each, as opposed to one group of 66 participants if only one age group had been assessed longitudinally) and a larger sample size may have reduced the need for the variables to be square root transformed in order to normalize the scores. While the transformation was successful in most cases, both 'lack' of metacognitive skills measures (Perseveration and Distraction) remained not normally distributed after the transformation. This led to the composite PD rate being used; while this produced some interesting results, it would have been valuable to examine the contributions that each variable made individually. However, arguably, a larger sample size would not have solved this problem, as by definition many children do not show any perseveration or distraction (and this is the reason they are unusual behaviors). Because of variables such as these, it was deemed most appropriate to conduct both parametric and nonparametric tests where possible, and to only report findings when the approaches corroborated one another (to reduce the risk of type I errors).

Despite the aforementioned limitations, many of which could be overcome, this developmentally sensitive approach to assessing metacognitive skills could prove extremely fruitful when used in a multitude of other designs, such as intervention or longitudinal, and settings, such as when children are working alone, in groups or in a classroom.

**Acknowledgments** This work was funded by an ESRC PhD studentship grant awarded to D.B. Thank you to Dr Sue Bingham and Charlotte Bargas for second coding.

## References

- Alexander, J. M., Carr, M., & Schwanenflugel, P. J. (1995). Development of metacognition in gifted children: directions for future research. *Developmental Review*, 15, 1–37.
- Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis* (2nd ed.). Cambridge: Cambridge University Press.

- Barcelo, F., & Knight, H. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, *40*, 349–356.
- Bekci, B., & Karakas, S. (2006). Metacognition and executive functions: related or different concepts. *International Journal of Psychophysiology*, *61*(3), 343.
- Butterfield, E. C., Nelson, T. O., & Peck, V. (1988). Developmental aspects of the feeling of knowing. *Developmental Psychology*, *24*(5), 654–663.
- Chevalier, N., & Blaye, A. (2008). Cognitive flexibility in preschoolers: the role of representation activation and maintenance. *Developmental Science*, *11*(3), 339–353.
- Crone, E. A., Ridderinkhof, K. R., Worm, M., Somsen, R. J. M., & van der Molen, M. W. (2004). Switching between spatial stimulus–response mappings: a developmental study of cognitive flexibility. *Developmental Science*, *7*(4), 443–455.
- De la Fuente, J., & Lozano, A. (2010). Assessing self-regulated learning in early childhood education: difficulties, needs, and prospects. *Psicothema*, *22*(2), 278–283.
- Deak, G. O., & Narasimham, G. (2003). Is perseveration caused by inhibition failure?—Evidence from preschool children's inferences about word meanings. *Journal of Experimental Child Psychology*, *86*(3), 194–222.
- Deloache, J. S., & Brown, A. (1987). The early emergence of planning skills in children. In J. Bruner & H. Haste (Eds.), *Making sense: The child's construction of the world* (pp. 108–130). London: Methuen & Co. Ltd.
- Devilly, G. J. (2005). ClinTools Software for Windows: Version 4. [www.clintools.com](http://www.clintools.com): Psytek Ltd.
- Eme, E., Puustinen, M., & Coutelet, B. (2006). Individual and developmental differences in reading monitoring: when and how do children evaluate their comprehension? *European Journal of Psychology of Education*, *21*(1), 91–115.
- Field, A. (2009). *Discovering statistics using SPSS*. London: Sage.
- Flavell, J. H., Beech, D. R., & Chinsky, J. M. (1966). Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development*, *37*, 283–299.
- Flavell, J. H., Green, F. L., Flavell, E. R., & Grossman, J. B. (1997). The development of children's knowledge about inner speech. *Child Development*, *68*, 39–47.
- Hacker, D. J., Dunlosky, J., & Graesser, A. C. (1998). *Metacognition in educational theory and practice*. Mahwah: Erlbaum.
- Karmiloff-Smith, A. (1979). Problem solving processes in children's construction and representations of closed railway circuits. *Archives de Psychologie*, *17*, 33–59.
- Keeney, T. J., Cannizzo, S. R., & Flavell, J. H. (1967). Spontaneous and induced verbal rehearsal in a recall task. *Child Development*, *38*(4), 953–966.
- Krebs, S. S., & Roebbers, C. M. (2010). Children's strategic regulation, metacognitive monitoring, and control processes during test taking. *British Journal of Educational Psychology*, *80*(3), 325–340.
- Lambert, B. (2001). Metacognitive problem solving in preschoolers. *Australian Journal of Early Childhood*, *26*(3), 24–30.
- Larkin, S. (2000). *How can we discern metacognition in year one children from interactions between students and teacher*. Paper presented at the ESRC Teaching and Learning Research Programme Conference.
- Molenaar, I., van Bostel, C. A. M., & Slegers, P. J. C. (2010). The effects of scaffolding metacognitive activities in small groups. *Computers in Human Behavior*, *26*, 1727–1738.
- Morton, J. B., & Munakata, Y. (2002). Active versus latent representations: a neural network model of perseveration, dissociation and decalage. *Developmental Psychobiology*, *40*(3), 255–265.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26). New York: Academic.
- Perrotin, A., Belleville, S., & Isingrini, M. (2007). Metamemory monitoring in mild cognitive impairment: evidence of a less accurate episodic feeling-of-knowing. *Neuropsychologia*, *45*, 2811–2826.
- Perrotin, A., Tournelle, L., & Isingrini, M. (2008). Executive functioning and memory as potential mediators of the episodic feeling-of-knowing accuracy. *Brain and Cognition*, *67*, 76–87.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In P. R. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Handbook of self-regulation* doi:<http://lib.myilibrary.com?ID=92724>.
- Puustinen, M. (1998). Help-seeking behavior in a problem-solving situation: development of self-regulation. *European Journal of Psychology of Education*, *13*(2), 271–282.
- Roderer, T., & Roebbers, C. M. (2010). Explicit and implicit confidence judgments and developmental differences in metamemory: an eye-tracking approach. *Metacognition and Learning*, *5*, 229–250.
- Roebbers, C. M., Schmid, C., & Roderer, T. (2009). Metacognitive monitoring and control processes involved in primary school children's test performance. *British Journal of Educational Psychology*, *79*, 749–767.
- Sangster, C. (2010). *The role of self-regulation in motor learning: exploring the self regulated motor performance of children with developmental coordination disorder*. PhD thesis, University of Cambridge, Cambridge.
- Schneider, W., & Pressley, M. (1997). *Memory development between two and twenty* (2nd ed.). Mahwah: Erlbaum.

- Schneider, W., Vise, M., Lockl, K., & Nelson, T. O. (2000). Developmental trends in children's memory monitoring: evidence from a judgment-of-learning task. *Cognitive Development, 15*, 115–134.
- Schneider, W., Knopf, M., & Sodian, B. (2009). Verbal memory development from early childhood to early adulthood. In W. Schneider & M. Bullock (Eds.), *Human development from early childhood to early adulthood*. New York: Psychology Press.
- Souchay, C., & Isingrini, M. (2004). Age related differences in metacognitive control: role of executive functioning. *Brain and Cognition, 56*, 89–99.
- Sperling, R. A., Walls, R. T., & Hill, L. A. (2000). Early relationships among self-regulatory constructs: theory of mind and preschool children's problem solving. *Child Study Journal, 30*(4), 233–251.
- Taconnat, L., Raz, N., Tocze, C., Bouazzaoui, B., Sauzeon, H., Fay, S., et al. (2009). Ageing and organisation strategies in free recall: the role of cognitive flexibility. *European Journal of Cognitive Psychology, 21*(2–3), 347–365.
- van Hout-Wolters, B., Simons, R., & Volet, S. (2000). Active learning: Self-directed learning and independent work. In R. Simons, J. van der Linden, & T. Duffy (Eds.), *New learning* (p. 16). Dordrecht: Kluwer.
- van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic.
- Veenman, M. V. J., Van Hout-Wolters, B. H. A. M., & Afflerbach, P. (2006). Metacognition and learning: conceptual and methodological considerations. *Metacognition and Learning, 1*, 3–14.
- Wellman, H. M. (1977). Preschoolers understanding of memory-relevant variables. *Child Development, 48*(4), 1720–1723.
- Whitebread, D., Anderson, H., Coltman, P., Page, C., Pino Pasternak, D., & Mehta, S. (2005). Developing independent learning in the early years. *Education, 3–13*(33), 40–50.
- Whitebread, D., Coltman, P., Pino Pasternak, D., Sangster, C., Grau, V., Bingham, S., et al. (2009). The development of two observational tools for assessing metacognition and self-regulated learning in young children. *Metacognition and Learning, 4*(1), 63–85.
- Whitebread, D., Pino Pasternak, D., Bryce, D., Vuillier, L., Buettner, G., Perels, F., et al. (2011). *Identifying and measuring self-regulation in young children: the development and validation of the CHILD observational instrument*. Paper presented at the EARLI Biennial Conference, University of Exeter, UK.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 279–306). Hillsdale: Erlbaum.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In P. R. Pintrich, M. Boekaerts, & M. Seidner (Eds.), *Handbook of self-regulation*. Orlando: Academic.
- Winsler, A., & Naglieri, J. (2003). Overt and covert verbal problem-solving strategies: developmental trends in use, awareness, and relations with task performance in children aged 5 to 17. *Child Development, 74* (3), 659–678.
- Winsler, A., Fernyhough, C., & Montero, I. (Eds.). (2009). *Private speech, executive functioning, and the development of verbal self-regulation*. Cambridge: Cambridge University Press.
- Zelazo, P. D., Muller, U., Frye, D., & Marcovitch, S. (2003). The development of executive function in early childhood. *Monographs of the Society for Research in Child Development, 68*(3), 1–137.
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In P. R. Pintrich, M. Boekaerts, & M. Zeidner (Eds.), *Handbook of self-regulation* doi:<http://lib.mylibrary.com?ID=92724>.
- Zimmerman, B. J. (2001). Theories of self-regulated learning—an overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 191–226). NJ: Erlbaum.