

UC Irvine

UC Irvine Previously Published Works

Title

Metacognitive monitoring in Kenyan, Chinese, and U.S. children.

Permalink

<https://escholarship.org/uc/item/6x00347b>

Journal

J Exp Child Psychol, 275

ISSN

0022-0965

Authors

Buehler, Florian Jonas

Lewis, Laura Simone

Zhang, Zhen

et al.

Publication Date

2026-10-19

DOI

10.1016/j.jecp.2026.106691

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Metacognitive monitoring in Kenyan, Chinese, and U.S. children

Florian Jonas Buehler^{a, ID}, Laura Simone Lewis^{a, b, ID}, Zhen Zhang^{c, d, *, ID},
Hanna Schleihau^{e, ID}, Yufei Zong^{c, d, f, ID}, Esther Herrmann^{g, ID}, Henriette Zeidler^{h, 1, ID},
Jan Maximilian Engelmann^{a, 1, ID}

^a Department of Psychology, University of California Berkeley, Berkeley, CA, USA

^b Depts. of Psychological and Brain Sciences and Anthropology, University of California Santa Barbara, Santa Barbara, CA, USA

^c CAS Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

^d Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

^e Department of Psychology, Utrecht University, Utrecht, Netherlands

^f Department of Cognitive Sciences, University of California Irvine, Irvine, CA, USA

^g School of Psychology, Sport and Health Sciences, University of Portsmouth, Portsmouth, UK

^h Sapienza University of Rome, Rome, Italy

ARTICLE INFO

Keywords:

Metacognition
Monitoring behavior
Development
Cross-cultural
Problem-solving
Kenya
China
US

ABSTRACT

Theories of cognitive development propose that cultural values shape metacognitive skills. However, cross-cultural research on metacognition remains limited and often relies on verbal assessments that may lack validity across cultural contexts. The present study examined metacognitive monitoring in Kenyan, Chinese, and U.S. children ($N = 95$; 6–10-year-olds; 46% female) using a non-verbal problem-solving task in which children assembled objects according to a model. Across all three cultural contexts, children exhibited various monitoring behaviors (e.g., checking the model), and the frequency of these behaviors increased with age. Cross-cultural differences emerged in the specific monitoring strategies children used: Chinese children engaged in more thorough monitoring than Kenyan and U.S. children. These findings highlight both culturally-recurrent and culturally-variable patterns in children's metacognitive monitoring.

Introduction

Metacognitive monitoring –the ability to evaluate ongoing cognitive processes (Nelson & Narens, 1990)- is a crucial milestone of cognitive development. Monitoring processes serve to retrieve relevant information and are the basis for guiding metacognitive control processes. The interaction between monitoring and control is crucial for regulating cognitive processes and associated behaviors during problem-solving. For instance, while working on a puzzle, a child may notice that they are repeatedly trying pieces that do not fit and feel unsure how to proceed (monitoring). In response, they may decide to first identify the corner pieces or sort pieces by color (control). In the long-term, accurate monitoring contributes to self-regulated learning and academic achievement (Dunlosky & Metcalfe, 2009; Roebers, 2017).

Children's monitoring develops from an early age. Preschool children show first signs of competent monitoring: Three-year-old children report higher confidence in correct than incorrect responses in perceptual tasks (Coughlin et al., 2015; Gonzales et al.,

* Corresponding author.

E-mail address: zhangzhen@psych.ac.cn (Z. Zhang).

¹ Shared Senior Authors.

2022; Lyons & Ghetti, 2011). By age four, children express greater confidence in accurate than inaccurate memories (Destan et al., 2014; Hembacher & Ghetti, 2014). Metacognitive monitoring becomes more robust with age (see the *meta*-analysis by Xia and colleagues, 2023a). Although young children show a well-replicated tendency to be overconfident (i.e., overestimating their performance) across different domains, this bias decreases with age (Xia et al., 2023a). Beyond confidence judgments, metacognitive monitoring is also reflected in children's behavior. For example, when constructing objects based on a model, 5 to 7-year-olds increasingly engage in detailed checking behavior, such as closely inspecting the model (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; 2026). Overall, metacognitive monitoring becomes more accurate, less biased, and more prevalent during the primary school years (for a review, see Roebbers, 2017).

Research on metacognitive development is almost exclusively based on data from children in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) populations (see Henrich et al., 2010). The WEIRD sampling bias limits the generalizability of findings and our understanding of the role of cultural context in metacognitive development (Arnett, 2008; Nielsen et al., 2017; Thalmayer et al., 2021). For instance, in the *meta*-analysis by Xia et al. (2023a), the effect of cultural context on overestimation could not be investigated due to a lack of studies involving non-WEIRD samples. It is thus unclear whether the developmental pattern reported in the previous paragraph generalizes to other cultural contexts. Which aspects of metacognitive development are culturally-recurrent, and which may vary with cultural context, remains unknown.

Recent sociocultural theories of metacognition underscore the need for cross-cultural research (Heyes et al., 2020; Langenhoff et al., 2025). The *cultural origins hypothesis* suggests that metacognition is transmitted and acquired through cultural learning (Heyes et al., 2020). Cultural learning describes social interactions between metacognitively skilled and less skilled individuals, such as adults and children. Through these interactions, metacognitive practices are transmitted via language, imitation, and teaching. In line with the *cultural origins hypothesis*, parents' and teachers' use of metacognitive language relates to children's metacognitive development (Carr et al., 1989; Coffman et al., 2008; Grammer et al., 2013; Gardier & Geurten, 2025). For example, children exposed to teachers using high memory-relevant instruction employed more sophisticated strategies in a memory task than children receiving low memory-relevant instruction (Grammer et al., 2013). Cultural values and beliefs may also affect metacognitive development. In some U.S. and European cultural contexts, for example, children's overconfidence (e.g., Xia et al., 2023a) may partly reflect parenting practices that emphasize confidence, praise, and standing out (e.g., Clegg et al., 2021; Xia et al., 2022). A key prediction of the *cultural origins hypothesis* is that metacognitive development should exhibit at least some degree of cross-cultural variation.

The number of previous cross-cultural studies on metacognitive development is very limited. Sunae Kim et al. (2021) studied the metacognitive performance of German and Japanese 3- to 5-year-old children in a paired-associates memory task. While children from both populations demonstrated monitoring, being more confident in correct than in incorrect memories, Japanese children demonstrated more accurate metacognitive control. When they were asked to sort their responses into a box for correct memories and a box for incorrect memories (and only correct memories in the correct memory box were rewarded), Japanese children were more likely than German children to select a correct memory and not select an incorrect memory for a reward. Sunae Kim et al. (2021) suggest that these differences might be grounded in cultural differences in individualism. In the less individualistic Japanese context, children may be more motivated to perform well when being evaluated by others than in the more individualistic German context. Hence, Japanese children may have been more eager to detect correct and incorrect memories in their sorting decisions. In addition, across two studies, Xia and colleagues (2022, 2023b) investigated overconfidence in Dutch and Chinese 4- to 5-year-old children. Children from both cultural contexts were similarly overconfident in their estimates of how far they could throw a ball and how many picture cards they could remember (Xia et al., 2022, 2023b). Children were also asked to estimate a peer's performance on the memory and ball-throwing tasks. Interestingly, Chinese children tended to overestimate a peer's performance more than their own, whereas Dutch children overestimated their own and a peer's performance to a similar extent (Xia et al., 2022). Xia et al. (2022) suggest that, for Chinese children, positively evaluating a peer may be culturally desirable and reflect well on themselves.

Previous cross-cultural studies have predominantly relied on verbal assessments of metacognition (Sunae Kim et al., 2021; Xia et al., 2022, 2023b). For example, children were asked by an experimenter to rate their confidence in response to questions such as, "How sure are you that your answer is correct?" or "How many picture cards can you remember?" (Sunae Kim et al., 2021; Xi et al., 2022). While verbal assessments have revealed important insights into children's developing metacognition, explicit metacognitive questions may not be equally common across cultural contexts. Moreover, cultural norms may shape experimenter-child interactions. For example, children in China and Kenya are generally encouraged to be modest in their interactions with adults (Binah-Pollak, 2014; Kinoti, 2010; Wu et al., 2002; Xu & Hamamura, 2014), whereas overt expressions of confidence are more culturally valued in the United States (Merkin, Taras, & Steel, 2014). Consistent with these cultural norms, Chinese students are less inclined than their American peers to present favorable self-assessments (Y. H. Kim et al., 2010). Thus, cultural differences in metacognition may reflect norms regarding confidence reporting (i.e., what individuals consider culturally appropriate to report) rather than genuine differences in metacognitive processes (i.e., what individuals actually think about their mental states).

Observations of children's behavior in problem-solving tasks represent an important and complementary alternative to verbal assessments (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; 2026; Eberhart et al., 2025; Marulis & Nelson, 2021). Metacognitive monitoring can be expressed behaviorally and investigated without recourse to complex metacognitive language. For instance, in the Train Track task by Bryce and Whitebread (2012), children assembled various toy train tracks according to a printed plan. Children engaged in metacognitive monitoring by checking their own construction or the plan. That is, they glanced back at the plan they were working on. Looking at the plan can support metacognitive monitoring by helping children update their representation of the goal state and compare it with their ongoing construction. This comparison may reveal a discrepancy or uncertainty, which can subsequently inform control processes, such as changing strategy or selecting the next piece to try. Behavioral measures may be less affected by culturally normative responding than verbal assessments because they can be observed without direct communication between the

child and experimenter.

While behavioral and verbal monitoring assessments are both based on Nelson and Narens (1990) framework of metacognitive processes, they capture slightly different aspects. Behavioral observations do not require participants to explicitly verbalize their metacognitive states and may therefore capture more implicit aspects of metacognition, whereas verbal assessments provide greater access to explicit metacognitive representations. For instance, glancing at the plan may be a relatively automated behavior that is not accompanied by a conscious verbal representation, such as, “I am not sure which object comes next.” This distinction may also help explain why behavioral and verbal measures of metacognition have been only loosely related in recent research (Eberhart et al., 2025). Behavioral observations may be particularly useful for cross-cultural research because they require little verbal interaction between the experimenter and participant during the assessment. Moreover, problem-solving tasks allow researchers to observe spontaneously occurring metacognitive monitoring processes in a potentially more ecologically valid setting.

The present study

To broaden our understanding of how cultural context shapes metacognition, we investigated the development of metacognitive monitoring in Kenyan, Chinese, and U.S. children. These cultural contexts differ in individualism, a dimension that has been suggested to be theoretically relevant for metacognition (Ordin et al., 2024; Sunae Kim et al., 2021; van der Plas et al., 2022). Contexts with low individualism typically value modesty over standing out or appearing overly confident. Consequently, less individualistic contexts may foster lower levels of overconfidence to meet social expectations around modesty (Ordin et al., 2024; van der Plas et al., 2022). In contrast, in more individualistic contexts, positively biased metacognition (i.e., overconfidence) may be beneficial for presenting oneself as a confident and competent member of society (e.g., van der Plas et al., 2022). Moreover, less individualistic contexts may promote thorough monitoring behavior because individuals are motivated to perform well when their actions are evaluated by others (e.g., Sunae Kim et al., 2021). The motivation to perform well may also increase sensitivity to internal signals of uncertainty, thereby promoting more thorough checking behavior, such as inspecting the plan. Based on assessments of cultural values (The Culture Factor, n.d.), individualism is lowest in the Kenyan context, higher in the Chinese context, and highest in the U.S. context. These differences make the three cultural contexts particularly well-suited for examining cultural variation in metacognitive monitoring.



Fig. 1. The Monkey Track Task in the three cultural contexts. Note. On the left is the participant's mat with 21 target objects and 21 distractor objects. On the right is the model with the 21 target objects. a) Kenya. The child glances at the model. b) U.S.. The child inspects the model. c) China. The child compares an object with the model.

We tested 6 to 10-year-old children. We chose this age range given previous findings that U.S. and European children's metacognition develops significantly throughout this age window (Roebbers, 2017; Xia et al., 2023a). We assessed metacognitive monitoring in the Monkey Track task (see Fig. 1), a behavioral problem-solving task similar to those used in previous studies with U.S. and European children (Bryce & Whitebread, 2012; Buehler & Oeri, 2024, 2026; Eberhart et al., 2025; Marulis & Nelson, 2021). In the Monkey Track task, children assemble everyday objects (e.g., pens, cups, bottles) to build a track according to a model. We video-recorded and analyzed the occurrences and duration of children's metacognitive monitoring behaviors (e.g., glancing at the model, inspecting the model, comparing objects with the model). While our focus was on nonverbal assessments, we also included a verbal measurement of metacognitive monitoring to compare our findings to previous cross-cultural studies. Similar to Xia et al. (2022, 2023b), we asked children to evaluate their overall performance after completing the Monkey Track task. We calculated children's metacognitive bias as the difference between estimated and actual performance (e.g., Schraw, 2009). A positive score indicates overconfidence, while a negative score indicates underconfidence.

Based on previous research and theoretical accounts linking individualism and modesty to metacognition, we speculated that Kenyan and Chinese children may demonstrate more advanced metacognitive abilities than U.S. children (Ordin et al., 2024; Sunae Kim et al., 2021; van der Plas et al., 2022). We speculated that Kenyan and Chinese participants may exhibit more frequent and prolonged monitoring behavior and less overconfidence than U.S. participants. We refrained from formulating stronger hypotheses because the existing empirical evidence is limited, current theoretical accounts remain vague, and, to the best of our knowledge, this is the first cross-cultural study on metacognitive behavior. Moreover, we predicted that monitoring behavior would become more frequent and prolonged with age, while overconfidence would decline (Bryce & Whitebread, 2012; Buehler & Oeri, 2024, Xia et al., 2023).

Method

Participants

The final sample consisted of 32 Kenyan ($M_{age} = 7.94$ years; $SD = 1.21$; range [6.00; 9.83 years]; 50% female), 36 Chinese ($M_{age} = 7.90$ years; $SD = 1.12$; range [6.08; 9.67 years]; 44% female), and 27 U.S. children ($M_{age} = 8.07$ years; $SD = 1.23$; range [5.63; 9.63 years]; 44% female). We based our sample size on studies that used behavioral observations to assess children's metacognitive monitoring, which typically included 30–35 participants per investigated group (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; Marulis & Nelson, 2021). We tested 7 additional U.S. children who had to be excluded due to experimenter errors ($n = 2$), technical issues with the camera ($n = 1$), or external distractions ($n = 4$). Kenyan children lived in the greater Nanyuki area, in the Mount Kenya region. Chinese children lived in Beijing. U.S. children mainly lived in the San Francisco Bay Area. We provide more detailed information on our three samples next. We collected data on children's ethnicity, socioeconomic background, and number of siblings with parental questionnaires in China and the U.S.. In Kenya, the teachers provided information on ethnicity and the number of siblings because we were unable to directly ask participants' parents. Data on annual household income could not be assessed in Kenya.

Kenya – Kikuyu

Kenyan children were from Kikuyu families the surroundings of Nanyuki, a town with ~ 70,000 inhabitants in the central part of the country (Kenya National Bureau of Statistics, 2019). Kikuyu form the largest ethnic group in Kenya and are influential in business and politics across many regions (Middleton & Kershaw, 2017). Traditionally, they have been small-scale farmers, cultivating crops and practicing animal husbandry for their subsistence (Whiting, 1996). Children in our sample came from more traditional rural households with low socioeconomic status (SES). In such settings, children typically grow up in large families with siblings, cousins, and friends of differing age groups and have few physical possessions (Price, 1996). Children in our sample had, on average, 2.78 siblings. Although Kenya is experiencing rapid societal changes, modesty, obedience, and respect for elders are still important values in Kikuyu culture, and children are expected to listen to and obey their parents and teachers (Kinoti, 2010). Praise is less common than in the U.S. (Shanley et al., 2024; Whiting 1996). From the age of five, almost all children attend school. Kikuyu children were tested by Kenyan research assistants at their local school. Tests were conducted in Swahili, the language of instruction at school. Kikuyu children were tested in September 2023.

China – Beijing

Chinese children came from Beijing. The tested children belonged mainly to the Han ethnic group (94%) and a minority to the Hui ethnic group (6%). Chinese children were mostly from middle to high-income households as indicated by the annual household incomes: 17% > 700'000 RMB, 31% > 450'000 RMB, 33% > 250'000 RMB, 17% < 250'000, 3% did not report their income. Many children in our sample grew up in one-child families (*Mean number of siblings* = 0.56), which is typical for Chinese families (Binah-Pollak, 2014). Even though individualistic values (e.g., materialism, successful career) are on the rise in modern Chinese society, traditional values are still perceived as important (Sun & Ryder, 2016; Xu & Hamamura, 2014). Children are encouraged to be modest and obedient (Binah-Pollak, 2014; Wu et al., 2002; Xu & Hamamura, 2014). Especially in public situations, it is socially more favorable to give modest self-evaluations than, for instance, in the U.S. (Y. H. Kim et al., 2010). Chinese parents are more likely to praise children for behaviors that reflect adherence to cultural expectations than praise to reinforce autonomy and individual achievement (Wang et al., 2008). Almost all children attend school from the age of six years on (Chen & Park, 2021). Children were tested by Chinese research assistants in the lab space. Testings were conducted in Mandarin. Chinese children were tested in July and August 2023.

U.S. - San Francisco Bay Area

U.S. children were from the San Francisco Bay Area. They were predominantly of two or more ethnicities (23%), 22% White, 15% Asian, 11% Hispanic, and 30% of parents did not provide information on their children's ethnicity. U.S. children were mostly from middle to high-income households, as indicated by the annual household incomes: 37% >200'000 USD, 7% >160'000 USD, 8% >120'000 USD, 4% >100'000 USD, 8% < 100'000 USD, 37% did not report their income. The San Francisco Bay Area is one of the wealthiest regions in the U.S. (United States Census Bureau, 2023). Children in the U.S. sample grew up with one sibling (*Mean number of siblings* = 0.94) and their parents. Other family members are typically not part of children's daily lives because they live elsewhere in the country. U.S. children grow up with many toys and their own possessions. Most young U.S. children attend nurseries and preschools, and school is mandatory from age 6 onward (California Department of Education, 2025). Children's autonomy and self-enhancement are emphasized by parents and in education from an early age (Keller, 2012; Sedikides et al., 2015). Moreover, children are often praised for their efforts and achievements (Brummelman et al., 2017; Clegg et al., 2021; Gunderson et al., 2013). U.S. children were tested in English by U.S. research assistants in the local zoo, science museum, and local summer camps. Testing in the U.S. took place in June and July 2023.

Procedure

The study was conducted in accordance with the Declaration of Helsinki. Ethical approval was obtained from local institutional review boards in Kenya (University of Eastern Africa Baraton), China (Institute of Psychology, Chinese Academy of Sciences), and the United States (University of California, Berkeley). We also received permission from Kenya's National Commission for Science, Technology, and Innovation to conduct the research. In China and the United States, we obtained parental consent before testing participating children. In Kenya, school directors and headteachers serve as children's official guardians at school. Therefore, we obtained guardian consent through them.

To ensure consistency in testing across the three study sites, the original English procedure was translated into Swahili by a native Kenyan speaker and then back-translated into English by another native speaker. The same process was followed with native Chinese speakers. Moreover, experimenters at all three study sites underwent extensive training. Participants were tested individually.

The Monkey Track Task

In the Monkey Track Task, participants were asked to assemble objects according to a given model (see Fig. 1). The task was presented in a child-friendly cover story: Participants were told that they were building a track for a little monkey that loves jumping over obstacles. To do this, they had to accurately recreate a model track by placing the correct objects in the right locations and orientations. To ensure similar task familiarity in Kenyan, Chinese, and U.S. children, we relied on everyday objects (e.g., cups, cardboard, pens, bottles). Piloting revealed that the Monkey Track worked well in all three cultural contexts and was suitable for 6–10-year-olds, yet challenging enough to observe metacognitive monitoring.

First, the participants completed a *familiarization trial* in which they received instructions and feedback to make sure they understood the task. For instance, when the participant placed an incorrect object, they were reminded that their track should match the model. The familiarization trial consisted of 9 target objects and 9 distractor objects. After completing the familiarization trial, participants moved on to the test trial. During the *test trial*, participants built the monkey track on their own and did not receive feedback. On the participant's mat were 21 target objects and 21 distractor objects. The 42 objects were randomly distributed on the participants' mat. The target objects were needed to copy the original monkey track. The distractor objects served to increase task difficulty and, consequently, the need for metacognitive skills. For example, when choosing between two perceptually similar objects (e.g., a small vs. a large cup), children may be more likely to engage in monitoring processes, such as checking whether they selected the correct object, than when discriminating between perceptually distinct objects (e.g., a cup vs. a pen). The target and distractor objects were similar in size, color, and shape (e.g., small vs big cup; red vs brown pen). A detailed description of the target and distractor objects is available in the Supporting Information (Fig. S1). Kenyan and Chinese participants were on their own in a room, and the experimenter waited outside the room. In the U.S., children cannot be left on their own in a room because of legal requirements. However, the experimenter explicitly did not observe the child during the test trial, leaving the setup and sitting in a chair facing away from the child. We left children on their own to reduce potential experimenter effects on children's behavior (e.g., Buehler & Oeri, 2024, 2026). When participants finished the task, they informed the experimenter. We video-recorded all test sessions using two cameras positioned at different angles, ensuring that children remained visible when they turned away from one camera. Additionally, we instructed children to build their Monkey Track on a mat, which helped ensure that the task remained within the cameras' field of view.

Coding scheme

We coded: monitoring behaviors (of three different varieties: glancing at the model, inspecting the model, and comparing an object with the model); task completion time; task performance; and metacognitive bias. For details, see below. The videos were coded by Kenyan, Chinese, and U.S. coders. Videos from all contexts were distributed across coders to prevent cultural bias in the coding process.

Metacognitive monitoring behaviors

We relied on the initial coding scheme for the Train Track by Bryce and Whitebread (2012) to observe and code metacognitive monitoring in the Monkey Track. In both the Train Track and the Monkey Track tasks, participants copy a model, and so we assumed

that we would observe similar monitoring behaviors. To create a detailed Monkey Track coding scheme, three coders first rated 10 randomly selected videos. Following Bryce and Whitebread's coding scheme (2012) and Nelson and Narens' (1990) metacognitive framework, raters coded checking behaviors that served to update the cognitive representation of the Monkey Track Task as indicators of metacognitive monitoring. The raters identified three monitoring behaviors that were reliably observable and relevant to task completion: glancing at the model, inspecting the model, and comparing objects with the model (see Fig. 1 for demonstrations of each behavior).

We were unable to reliably code all behaviors that have been coded in the Train Track Task (Bryce & Whitebread, 2012; Eberhardt et al., 2025). For instance, we did not code behaviors that occurred on the participant's mat (e.g., checking one's own construction or searching for an object), as these behaviors could not be reliably distinguished from one another. We suspect that this was due to the task setup, in which all objects were placed on the participant's mat. For instance, checking the construction and searching for an object could look very similar. Nevertheless, children engaged in more than 8 monitoring behaviors per minute, which is similar to previous studies (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; Eberhardt et al., 2025).

The final coding scheme is presented in Table 1, and a more detailed version is provided in the Supporting Information (Table S1). As suggested by Buehler and Oeri (2024), we coded the occurrences and duration of monitoring behaviors for each participant. While occurrences and duration reflect the same underlying monitoring process, they differ in their sensitivity to specific behaviors. For example, glances are frequent (>7/minute), but each instance is brief (~1.5 s), such that meaningful variation is better captured by occurrences than by duration. In contrast, inspecting and comparing objects typically takes more time per instance (>4 s), but occurs less frequently (<0.5/minute), and is therefore better captured by duration than by occurrences. We coded the videos with ELAN version 5.9 (Max Planck Institute for Psycholinguistics The Language Archive, 2024). We double-coded 31 (33%) of the videos. Interrater reliabilities for glancing, inspecting, and comparing were excellent, ranging from ICC = .94 to .98 for occurrences and durations. ICC scores were calculated for each child's sum score per behavior. To account for differences in task completion times among children, we computed the per-minute occurrences and durations of monitoring behaviors. The monitoring behaviors are displayed in Fig. 1, and mean scores are presented in Table 2.

Task completion time

We computed task completion time as the difference between the end time and the start time. Start time was defined as the first instance of touching an object or the first observed monitoring behavior (typically glancing at the model) after the instructions were completed. We did not define the end of instructions as the start time because some children waited until the experimenter had left the room before beginning the task. The end time was defined as walking off camera or verbally indicating that they were done with the task. The start and end times for all subjects were defined by one coder.

Task performance

To assess task performance, we rewarded one point for each target object that was placed in the right order and orientation (tolerance of 45 degrees deviation) on the participants' mat. A score of 100% indicates that all 21 target objects were correctly placed on the participant's mat, in the correct order and orientation (see Table 1). Two raters independently computed all performance scores (ICC = .96). The raters solved disagreements by joint inspections of the respective videos.

Table 1
Monkey Track coding scheme.

Ratings	Example
<i>Monitoring behavior</i>	
Glancing at the model	The participant glances at the original model.
Inspecting the model	The participant walks over to the original model to get a closer look.
Comparing an object with the model	The participant takes an object over to the original model to compare it with an object in the original model.
Total monitoring	The sum score of glancing, inspecting, and comparing.
Task completion time	Time spent to complete the task.
<i>Metacognitive bias</i>	
Actual Performance	Correctly placed target objects on the participant's mat.
Estimated Performance	How many objects did you place correctly?
Metacognitive bias	Estimated performance – Actual performance

Note. A more detailed version of the coding scheme can be found in the Supporting Information (Table S1).

Table 2

Mean (SD) for metacognitive bias and monitoring behaviors overall, per cultural context, and per age group (median split).

Variable	Overall	Kenya	China	U.S.	Younger	Older
<i>N</i>	95	32	36	27	47	48
Metacognitive bias						
Actual performance [%]	88.86 (11.86)	81.25 (12.39)	95.37 (7.06)	89.20 (11.31)	86.03 (12.64)	91.57 (10.50)
Estimated performance [%]	90.96 (17.14)	89.84 (19.94)	92.36 (14.42)	90.38 (17.43)	93.48 (17.03)	88.54 (17.07)
Metacognitive bias [%]	2.10 (18.88)	8.59 (20.73)	−3.01 (17.02)	1.19 (17.19)	7.45 (17.69)	−3.03 (18.73)
Monitoring behavior						
Task completion time [sec]	285.22 (83.68)	298.57 (57.70)	313.06 (86)	232.3 (84.79)	288.26 (82.35)	282.25 (85.73)
Glancing [occ/min]	7.61 (2.39)	8.30 (2.59)	6.70 (1.68)	7.99 (2.62)	6.88 (2.45)	8.32 (2.11)
Glancing [dur / min]	10.96 (3.31)	11.05 (2.7)	10.32 (2.81)	11.68 (4.38)	10.85 (3.75)	11.06 (2.85)
Inspecting [occ/min]	0.47 (0.58)	0.24 (0.29)	0.89 (0.64)	0.19 (0.44)	0.47 (0.57)	0.47 (0.61)
Inspecting [dur / min]	1.95 (2.56)	0.83 (1.38)	3.82 (2.68)	0.80 (1.98)	1.77 (2.14)	2.13 (2.92)
Comparing [occ/min]	0.26 (0.33)	0.13 (0.28)	0.47 (0.30)	0.13 (0.26)	0.20 (0.28)	0.32 (0.36)
Comparing [dur / min]	1.33 (1.95)	0.57 (1.43)	2.56 (2.27)	0.59 (1.02)	0.86 (1.46)	1.79 (2.26)
Total behavior [occ/min]	8.34 (2.27)	8.66 (2.49)	8.07 (1.55)	8.32 (2.80)	7.55 (2.40)	9.11 (1.87)
Total behavior [dur / min]	14.24 (4.17)	12.46 (3.02)	16.71 (2.88)	13.08 (5.21)	13.49 (4.40)	14.99 (3.84)

Note. Metacognitive bias = Estimated performance – Actual performance. Occ / min = occurrences per minute, dur / min = seconds per minute.

Metacognitive bias

After completing the Monkey Track, we asked the participants on a 5-point Likert scale how many objects they thought they had placed correctly. Similar to previous research (e.g., Buehler et al., 2025), we used a smiley scale, ranging from “none of the objects were placed correctly” (depicted by a sad-looking smiley) to “all of the objects were placed correctly” (depicted by a happy-looking smiley; see Supporting Information Fig. S2). To obtain an estimate of task performance, we converted the 5-point scale to percentages (none = 0%, some = 25%, half = 50%, most = 75%, all = 100%). We then computed a metacognitive bias score, defined as the difference between estimated and actual task performance (Schraw, 2009). A positive score indicates overconfidence, a negative score indicates underconfidence and a score of 0 indicates perfect metacognitive calibration. See Table 2 for mean values.

Statistical analyses

We conducted statistical analyses in R (R Core Team, 2025; Version 4.4.3). We tested the effects of cultural context and age on overall occurrences (Model 1) and duration (Model 2) of metacognitive monitoring behavior, and on metacognitive bias (Model 3), using linear models with the base R function `lm()`. In all models, we included main effects and interaction terms of cultural context and age. While not being part of our primary hypotheses, the interaction term allows us to explore cross-cultural differences in metacognitive development. In Models 1 and 2, we also controlled for task performance because it is closely related to monitoring (e.g., Dunlosky et al., 2016; Fleming & Lau, 2014). Task performance may vary across cultural contexts and age groups, and thus, observed group differences in monitoring could partly reflect performance differences rather than differences in monitoring behavior itself. Controlling for task performance allows us to test whether cultural context and age effects in monitoring persist when task performance is held constant. However, we did not control for task performance in Model 3 because metacognitive bias (estimated performance – task performance) and task performance are mathematically coupled. We z-standardized age (in months) and task performance. For all analyses, we first determined the significance of the full model by comparing it to a null model using the base R function `anova()`. To test the significance of individual predictors, we used likelihood-ratio tests with the `drop1()` function from base R. We followed up on significant interactions or main effects for cultural context with Tukey corrected pairwise comparisons using the `emmeans` package (Lenth 2025; Version: 1.10.7). We further compared individual monitoring behaviors (e.g., glancing, inspecting, comparing) between cultural contexts. Because individual behaviors were right-skewed, we relied on non-parametric Kruskal-Wallis tests and followed up with Dunn tests with Holm-corrected p-values for pairwise comparisons using the `dunn.test` package (Dinno, 2024; Version: 1.3.6).

As part of exploratory analysis, we computed Spearman correlations between individual monitoring behaviors and task performance, and between overall monitoring behaviors and metacognitive bias using the `ppcor` package (Seongho Kim, 2015; Version 1.1). For all analyses, we used a significance threshold of 0.05 for p-values and reported trends between 0.05 and 0.10 to avoid arbitrary dichotomization of significance (McCulloch & Searle, 2000). The analyses were not preregistered. R scripts and dataframes can be accessed here: <https://osf.io/f8gw2/>.

Results

Descriptives

Descriptives for the Monkey Track are presented in Table 2. Kenyan, Chinese, and U.S. children showed metacognitive behaviors. Overall, participants engaged in $M = 8.34$ ($SD = 2.27$) monitoring behaviors per minute, lasting for $M = 14.24$ s ($SD = 4.17$) per minute. The most prevalent monitoring behavior was glancing at the model, followed by inspecting the model, and then comparing objects with the model.

Monkey Track completion time was highest for Chinese children ($M = 313.06$ s; $SD = 86.00$), followed by Kenyan ($M = 298.57$ s;

$SD = 57.70$), and then U.S. children ($M = 232.30$ s; $SD = 84.79$). Tukey-corrected Pairwise comparisons revealed that U.S. children completed the Monkey Track faster than Kenyan ($Mdiff = 66.30$, $t(92) = 3.28$, $p = .004$) and Chinese children ($Mdiff = 80.80$, $t(92) = 4.11$, $p < .001$); task completion time did not differ between Kenyan and Chinese children ($Mdiff = 14.50$, $t(92) = 0.77$, $p = .721$). Moreover, task completion time was similar in younger ($M = 288.26$; $SD = 85.730$) and older children ($M = 288.26$; $SD = 82.35$) ($Mdiff = 6.01$, $t(93) = 0.35$, $p = .728$). Task performance was highest among Chinese children (95.37%), followed by U.S. (89.20%) and Kenyan (81.25%) children. Tukey-corrected pairwise comparisons revealed that Kenyan children scored lower than Chinese ($Mdiff = 14.12$, $t(91) = 5.63$, $p < .001$) and U.S. children ($Mdiff = 7.94$, $t(91) = 2.91$, $p = .012$), and a trend for Chinese children to score higher than U.S. children ($Mdiff = 6.18$, $t(91) = 2.32$, $p = .058$). Task performance was also higher in older ($M = 91.57$, $SD = 10.50$) than in younger children ($M = 86.03$, $SD = 12.64$) ($Mdiff = 5.54$, $t(92) = 2.32$, $p = .023$).

Models 1 and 2: Predicting monitoring behavior by age and cultural context

We first tested the effects of cultural context, age, and the cultural context $\times$ age interaction on monitoring occurrences (Model 1) and duration (Model 2). We controlled for task performance in all models. To compute overall monitoring scores, we summed the per-minute occurrences and durations of three monitoring behaviors: glancing at the model, inspecting the model, and comparing objects.

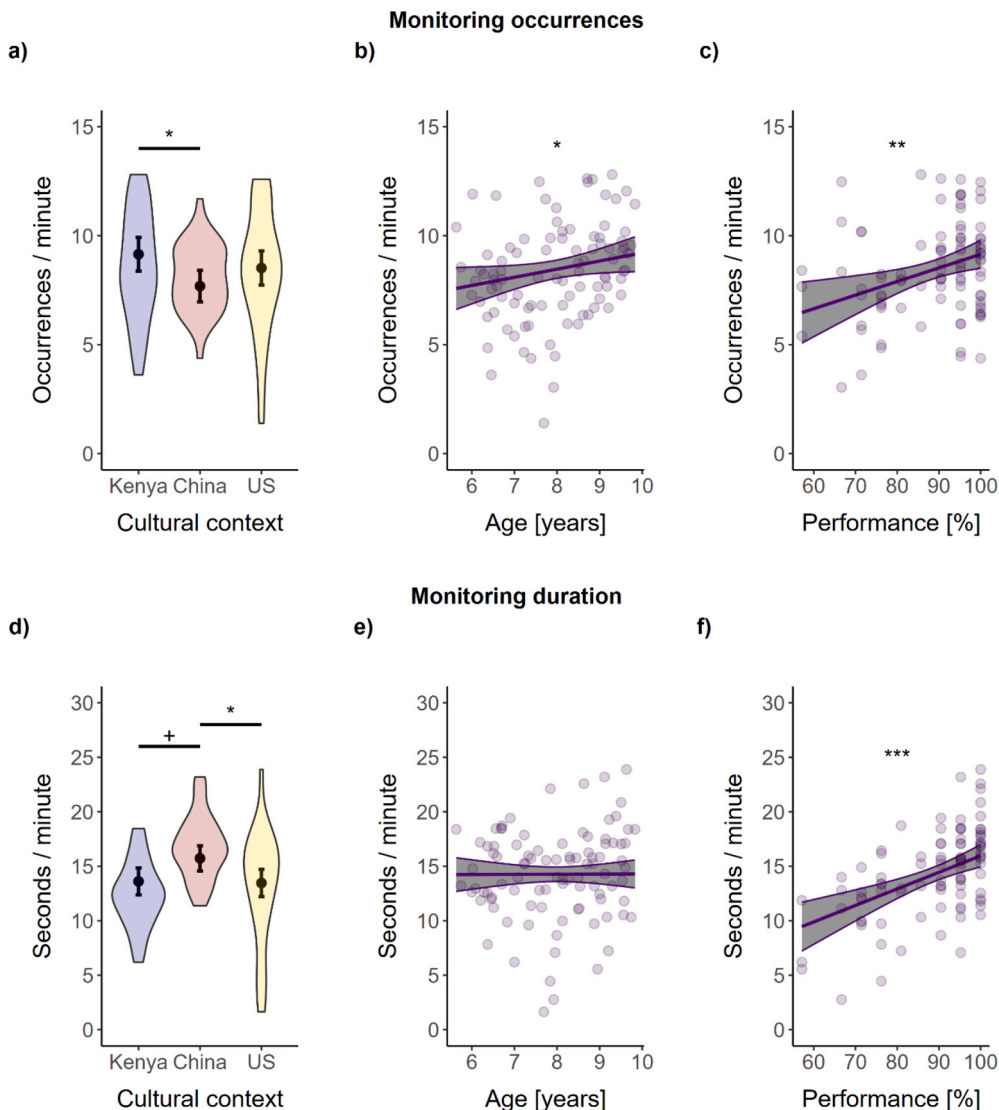


Fig. 2. Predictors of monitoring occurrences and monitoring duration. Note. Plots a, b, and c show the terms of the linear model: monitoring occurrences $\sim$ cultural context + age + performance. Plots d, e, and f show the terms of the linear model: monitoring duration $\sim$ cultural context + age + performance. Plots a and d show the predicted means and 95% confidence intervals by cultural context; plots b and e show them by age; and plots c and f show them by task performance. The small dots represent the observed data. * $p < .05$; ** $p < .01$; *** $p < .001$; + $p < .07$.

Model 1: For monitoring occurrences, the full model (monitoring occurrences ~ cultural context*age + task performance) explained the data better than the corresponding null model (monitoring occurrences ~ 1 + task performance), ($\chi^2(5) = 56.76$, $p = .016$). We found no cultural context*age interaction ($\chi^2(2) = 5.00$, $p = .517$), and therefore dropped the interaction in the reduced model. In the reduced model, we found a significant effect of cultural context ($\chi^2(2) = 26.46$, $p = .035$), age ($\chi^2(1) = 16.27$, $p = .041$, $\beta = 0.20$), and task performance ($\chi^2(1) = 34.34$, $p = .003$, $\beta = 0.34$). Tukey-corrected pairwise comparisons revealed that Kenyan children showed more monitoring occurrences than Chinese children ($Mdiff = 1.46$, $t(89) = 2.54$, $p = .034$; $d = 0.73$), but no differences between Kenyan and U.S. children ($Mdiff = 0.63$, $t(89) = 1.14$, $p = .495$; $d = 0.31$), nor between Chinese and U.S. children ($Mdiff = 0.83$, $t(89) = 1.55$, $p = .273$; $d = -0.41$). Contrary to our predictions, the frequency of monitoring behavior did not differ between Kenyan and U.S. children, nor between Chinese and U.S. children. However, Kenyan children monitored more frequently than Chinese children. Consistent with our expectations, monitoring frequency increased with age. Finally, monitoring occurrences increased with task performance (see Fig. 2a-c). We additionally explored the same models without controlling for task performance (see Supporting Information Results S1). These analyses confirmed that age predicted the occurrence of monitoring behaviors but revealed no effects of cultural context. This pattern indicates that the association between cultural context and monitoring occurrences is closely related to task performance and is not robust across statistical models. Cross-cultural differences should therefore be interpreted with caution.

Model 2: For monitoring duration, the full model (monitoring duration ~ cultural context*age + task performance) did not significantly differ from the null model (monitoring duration ~ 1 + task performance), ($\chi^2(5) = 94.23$, $p = .109$). However, the reduced model (monitoring duration ~ cultural context + age + task performance) explained the data significantly better than the null model, ($\chi^2(3) = 86.59$, $p = .039$). In the reduced model, we found a significant effect of cultural context ($\chi^2(2) = 84.43$, $p = .016$), no effect of age ($\chi^2(1) = 0.02$, $p = .962$, $\beta = 0.00$), and an effect of task performance ($\chi^2(1) = 202.56$, $p < .001$, $\beta = 0.45$). Tukey-corrected pairwise comparisons revealed that Chinese children showed higher monitoring duration than U.S. children ($Mdiff = 2.26$, $t(89) = 2.63$, $p = .027$; $d = 0.70$), a trend towards higher monitoring duration in Chinese than Kenyan children ($Mdiff = 2.11$, $t(89) = 2.29$, $p = .062$; $d = 0.66$), but no differences between Kenyan and U.S. children ($Mdiff = 0.15$, $t(89) = 0.17$, $p = .985$; $d = 0.05$). As predicted, monitoring duration was higher in Chinese than in U.S. children. Moreover, Chinese children tended to monitor for longer than Kenyan children. Contrary to expectations, the monitoring duration did not differ between Kenyan and U.S. children, nor did it increase with age. Finally, monitoring duration increased with task performance (see Fig. 2d-f). We additionally explored the same models without controlling for task performance, confirming the same pattern of results (see Supporting Information Results S2).

To better understand cross-cultural differences in the overall monitoring behaviors found in Models 1 and 2, we analyzed individual monitoring behaviors. Specifically, we compared the use of glancing at the model, inspecting the model, and comparing objects between Kenyan, Chinese, and U.S. participants. We relied on proportional monitoring scores. For each monitoring behavior, we calculated a proportional score by dividing its number of occurrences by the total number of occurrences of all monitoring behaviors, and its duration by the total duration of all monitoring behaviors (see Fig. 3).

Glancing occurrences (Kenya = 96%; China = 83%; U.S. = 96%) and duration (Kenya = 89%; China = 62%; U.S. = 89%) differed significantly between cultural contexts (occurrences: $\chi^2(2) = 39.51$, $p < .001$; duration: $\chi^2(2) = 39.86$, $p < .001$). Chinese children showed less glancing at the model than Kenyan ($p_{occurrences} < .001$; $p_{duration} < .001$), and U.S. children ($p_{occurrences} < .001$; $p_{duration} < .001$), while Kenyan and U.S. children did not differ ($p_{occurrences} = .209$; $p_{duration} = .283$). Inspecting occurrences (Kenya = 3%; China = 11%; U.S. = 2%) and duration (Kenya = 7%; China = 23%; U.S. = 6%) also differed significantly between cultural contexts (occurrences:

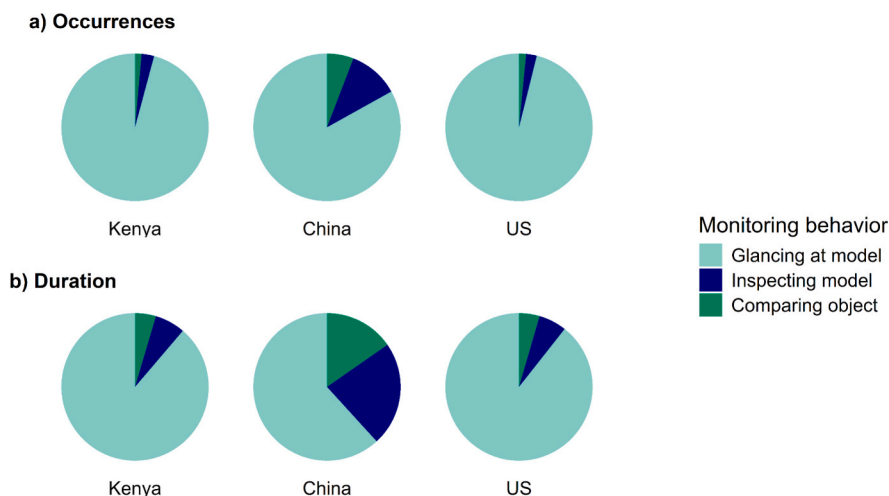


Fig. 3. Proportions of (a) occurrences and (b) duration of individual monitoring behavior per cultural context. Note. The displayed scores represent the relative use of each individual monitoring behavior as a proportion of all monitoring behaviors, presented separately for each cultural context. Glancing at the model is the most prevalent behavior in all three cultural contexts. Chinese children rely more on inspecting the model and comparing objects with the model than Kenyan and U.S. children.

$\chi^2(2) = 37.81, p < .001$; duration: $\chi^2(2) = 37.04, p < .001$). Chinese children inspected the model more than Kenyan children ($p_{\text{occurrences}} < .001$; $p_{\text{duration}} < .001$) and U.S. children ($p_{\text{occurrences}} < .001$; $p_{\text{duration}} < .001$), while Kenyan and U.S. children did not differ ($p_{\text{occurrences}} = .116$; $p_{\text{duration}} = .143$). Comparing objects occurrences (Kenya = 1%; China = 6%; U.S. = 2%) and duration (Kenya = 5%; China = 15%; U.S. = 5%) differed significantly between cultural contexts as well (occurrences: $\chi^2(2) = 35.07, p < .001$; duration: $\chi^2(2) = 31.76, p < .001$). Chinese children showed more comparisons of an object with the model than Kenyan ($p_{\text{occurrences}} < .001$; $p_{\text{duration}} < .001$) and U.S. children ($p_{\text{occurrences}} < .001$; $p_{\text{duration}} < .001$), while Kenyan and U.S. children did not differ ($p_{\text{occurrences}} = .364$; $p_{\text{duration}} = .310$).

In summary, Chinese children relied less on glancing at the model but more on inspecting the model and comparing objects with the model than Kenyan and U.S. children. Kenyan and U.S. children showed similar patterns of glancing at the model, inspecting the model, and comparing pieces with the model. We replicated these results with absolute values, meaning that Chinese children not only allocated a greater proportion of their monitoring to inspecting and comparing, but also engaged in these behaviors more frequently and longer (see Table 2 and Supporting Information Results S3).

Model 3: Predicting metacognitive bias by cultural context and age

We tested the effects of cultural context, age, and the cultural context $\times$ age interaction on metacognitive bias (Model 3). The full model (metacognitive bias $\sim$ cultural context*age) fit the data better than its corresponding null model (monitoring occurrences $\sim$ 1), $\chi^2(5) = 4887.10, p = .009$. We found no cultural context*age interaction ($\chi^2(2) = 156.80, p = .771$) and therefore dropped the interaction from the reduced model. Testing the reduced model, we found an effect of age ($\chi^2(1) = 2419.80, p = .006, \beta = -.27$), suggesting that metacognitive bias was lower in older children. Finally, we found an effect of cultural context $\chi^2(2) = 2348.80, p = .024$. Tukey-corrected pairwise comparisons revealed that metacognitive bias was larger in Kenyan than in Chinese children ($M_{\text{diff}} = 11.76, t(90) = 2.72, p = .021; d = 0.66$), but no difference between U.S. and Chinese ($M_{\text{diff}} = 5.00, t(90) = 1.09, p = .52; d = 0.28$), and between Kenyan and U.S. children ($M_{\text{diff}} = 6.76, t(99) = 1.44, p = .325; d = 0.38$). The mean scores of metacognitive bias in Table 2 suggest that Kenyan children were most overconfident.

However, when controlling for task performance in the models, the effects of cultural context and age on metacognitive bias no longer hold (see Supporting Information, Results S4). Although these models should be interpreted with caution, given the mathematical coupling between task performance and metacognitive bias (bias = estimated performance – task performance), they suggest that differences in metacognitive bias across cultural contexts and ages may reflect differences in task performance.

Exploratory analyses

To examine the relationship between specific metacognitive monitoring behaviors and task performance, we calculated Spearman correlations between each monitoring behavior (glancing, inspecting, and comparing) and task performance. Age-controlled Spearman correlations revealed that increased inspection of the model (occurrences: $\rho = .50, p < .001$; duration: $\rho = .51, p < .001$) and increased comparing of objects with the model (occurrences: $\rho = .41, p < .001$; duration: $\rho = .41, p < .001$) were significantly associated with higher task performance. In contrast, glancing at the model was not (occurrences: $\rho = -.06, p = .540$; duration: $\rho = .01, p = .960$). To better understand the nonsignificant relation between glancing and task performance, we divided the sample into high and low performers with a median split. Interestingly, glancing was positively associated with task performance for low performers (occurrences: $\rho = .16, p = .377$; duration: $\rho = .34, p = .024$), and negatively associated with task performance among high performers (occurrences: $\rho = -.32, p = .027$; duration: $\rho = -.37, p < .001$). This pattern suggests that the relation between glancing and task is inverted for low and high performers.

We also explored the relation of monitoring components by testing the association between monitoring behavior and metacognitive bias. Age-controlled Spearman correlations revealed that monitoring duration ($\rho = -.22, p = .033$), but not occurrences ($\rho = -.09, p = .377$), were related to metacognitive bias. However, performance-controlled Spearman correlations revealed no association between metacognitive bias and monitoring (occurrences: $\rho = -.07, p = .478$; duration: $\rho = .15, p = .149$). Taken together, these findings suggest that metacognitive behavior is not robustly associated with metacognitive bias after accounting for task performance.

Discussion

We investigated metacognitive monitoring among 6- to 10-year-old children from Kenya, China, and the United States using a non-verbal problem-solving task. Participants were asked to assemble objects according to a model, allowing us to observe three different monitoring behaviors: glancing at the model, inspecting the model, and comparing pieces with the model (see Fig. 1). We replicated previous research with U.S. and European children by demonstrating that children in three diverse cultural contexts exhibit monitoring behaviors (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; 2026; Marulis & Nelson, 2021). Moreover, across all cultures, monitoring frequency increased with age. Cultural differences emerged when zooming in on the monitoring behaviors: Chinese children monitored more extensively than Kenyan and U.S. children, both in terms of duration and type of behaviors – they were more likely to inspect the model and to actively compare objects. Kenyan and U.S. children showed similar patterns of monitoring behavior. This is the first cross-cultural study to explore metacognitive monitoring in a non-verbal context, shedding light on both culturally recurrent and culturally specific aspects of metacognition.

Metacognitive monitoring behaviors across cultural contexts

We first examined how cultural context and age influenced the frequency and duration of overall monitoring behaviors, controlling for task performance (see Fig. 2). Consistent with previous research (Bryce & Whitebread, 2012), we found that the rate of monitoring behavior increased with age, whereas the duration of monitoring behavior did not. As children develop, they may monitor their performance more efficiently. They engage in monitoring more often, but each instance is briefer. The control variable task performance revealed that children who monitored more often and longer also more accurately assembled the Monkey Track. This illustrates that the observed monitoring behaviors reflect metacognitive processes that improve children's ability to successfully complete the task.

Kenyan children monitored more frequently than Chinese children, while Chinese children monitored longer than Kenyan and U.S. children. Inspecting the individual metacognitive monitoring behaviors (Fig. 3) clarifies these mixed findings. Chinese children were more likely to inspect the model and to compare objects with the model, but less likely to glance at the model than Kenyan and U.S. children. A single instance of walking to the model for inspecting or comparing an object with the model takes more time than quickly glancing at the model. Hence, inspecting the model and comparing objects with the model is more apparent in the overall duration score, whereas glancing at the model is more visible in the overall occurrences score. In summary, Chinese children utilized different metacognitive behaviors than U.S. and Kenyan children.

Kenyan and U.S. children exhibited highly similar monitoring behaviors. They predominantly relied on glancing at the model instead of more detailed but time-consuming behaviors, such as inspecting the model or comparing pieces with the model. U.S. children may have prioritized speed and efficiency as they also completed the Monkey Track faster than Chinese and Kenyan children. Kenyan children may have been slightly less familiar with the task, as suggested by their lower task performance. Indeed, toys similar to the Monkey Track, such as Legos and Puzzles, are very popular in China and the U.S., but mostly unknown in Kenya. Quick and frequent glancing behavior may be the most adaptive and pragmatic approach in an unfamiliar task context because it reduces cognitive load through immediate and frequent feedback on task progress.

Metacognition may be shaped by cultural norms and practices (Heyes et al., 2020). For example, less individualistic contexts that emphasize modesty, may encourage children to closely monitor their cognitive processes (Sunae Kim et al., 2021; van der Plas et al., 2022). The current results are only partly in line with this interpretation. Individualism is lower in China than in the U.S. (The Culture Factor, n.d.), which may help explain why Chinese children exhibited more thorough monitoring behaviors than their U.S. counterparts. However, individualism is even lower in Kenya than in China (The Culture Factor, n.d.), but Kenyan and U.S. children do not differ in their monitoring behavior. These inconsistencies highlight the speculative nature of these interpretations. It is important to note that we did not directly measure individualism in the present study. Instead, we relied on descriptive indices assessed in the work context (The Culture Factor, n.d.). Future research would benefit from directly measuring individualism to better understand its relation to metacognitive behavior. Moreover, individualism may be too broad to capture the specific cultural mechanisms driving variation in metacognition (e.g., Voronov & Singer, 2002; Wong et al., 2018). Future studies should therefore also investigate more proximal cultural values and socialization practices.

Educational practices could be an interesting avenue for future research. For instance, Chinese parents and teachers typically emphasize diligence, task performance, and perfection, which may foster monitoring (Ng & Wei, 2020; Ran, 2001; Zhu & Chang, 2019). The ideal of perfection, deeply embedded in Confucian teachings, is pursued through continual self-improvement (Ng & Wei, 2020). To achieve perfection, Chinese parenting involves rigorous training characterized by constant monitoring and correction (Chao, 2000; Hou et al., 2021; Huntsinger & Jose, 2009). For instance, Chinese parents gave more instructions when assembling a Puzzle with their child than American parents (Hou et al., 2021). Importantly, this training is viewed as an expression of care and love for children (see Ng et al., 2007). The emphasis on performance and perfection aligns with the observed monitoring behaviors: inspecting the model and comparing objects with the model were positively correlated with task performance, while glancing at the model was not related to task performance in high-performing children. These findings suggest that Chinese children engage in more thorough metacognitive monitoring than their Kenyan and U.S. counterparts. We speculate that the focus on high performance, perfection, and obedience in Chinese education may encourage children to monitor their performance more closely to maximize task performance. Assessing educational characteristics in future cross-cultural research would strengthen such conclusions. For example, prior research in WEIRD contexts has examined parents and teachers' use of metacognitive language in relation to children's metacognition (Gardier & Geurten, 2025; Geurten & Léonard, 2023; Coffman et al., 2008; Grammer et al., 2013). Actively assessing parental and teacher practices would help illuminate the role of culturally embedded education strategies in shaping metacognitive skills.

Moreover, differences in persistence and motivation may also contribute to cross-cultural differences (Buehler et al., 2026; Marulis & Nelson, 2021; Yanaoka et al., 2024). For instance, task persistence and motivation have been associated with greater engagement in metacognitive behaviors (Buehler et al., 2026; Marulis & Nelson, 2021). Chinese children are typically socialized to follow adult requests closely (Binah-Pollak, 2014; Wu et al., 2002; Xu & Hamamura, 2014). Thus, they may be less likely to disengage from a task than U.S. children, whose autonomy and self-enhancement are more strongly emphasized by parents and teachers (Keller, 2012; Sedikides et al., 2015). In line with this speculation, Chinese children spent significantly more time on the monkey track task than U.S. children did. Altogether, these findings highlight the need to consider a broader range of motivational and behavioral factors when investigating cross-cultural differences in metacognitive behavior.

Overconfidence across cultural contexts

We also assessed children's metacognition verbally using a metacognitive bias score, which indicates the extent to which children's

estimated performance matches their actual performance. In line with a recent meta-analysis by Xia et al. (2023a), children overestimated their performance, but overestimation decreased with age. Interestingly, overestimation was significantly higher among Kenyan than among Chinese children, although it did not differ between the U.S. and Chinese samples or between the U.S. and Kenyan samples. Previous research suggests that East Asian samples tend to be more metacognitively sensitive. For instance, compared with Western samples, they report lower confidence in incorrect trials and are more likely to incorporate their confidence judgments into metacognitive control decisions (Sunae Kim et al., 2021; van der Plas et al., 2022). A popular explanation for an East Asian advantage in metacognitive judgments is that cultural practices emphasize modesty and, therefore, prevent overconfidence (Ordin et al., 2024; Sunae Kim et al., 2022; van der Plas et al., 2022). However, given that Kenyan children are also raised to be modest members of society, this explanation does not readily account for the difference between Kenyan and Chinese children, suggesting that other factors may be at play.

Importantly, when controlling for task performance, cultural context, and age were no longer related to metacognitive bias. Task performance was highest in the Chinese sample and lowest in the Kenyan sample (Kenya: 81.25%; China: 95.37%; U.S.: 89.20%), and also higher in older (91.57%) than in younger (86.03%) children. Thus, differences in task performance may contribute to the observed cross-cultural differences and age effects in overconfidence. Therefore, the cross-cultural differences and age effects should be interpreted with caution. Future research could address this issue by matching task performance across cultural contexts and age groups using ability-adapted tasks.

Another important point to consider is that the overconfidence level in our sample (2.10%) was well below the approximately 30% reported in a recent meta-analysis (Xia et al., 2023a). The low levels of overconfidence may have limited cross-cultural and age-related variability. One reason for the low overconfidence in our sample might be that we assessed children's performance estimates retrospectively (i.e., after they had completed the Monkey Track Task), while Xia and colleagues focused on prospective performance estimates. Retrospective judgments tend to be more accurate than prospective judgments (Cottini et al., 2021; Lipko et al., 2009; Stelz & Altgassen, 2024). In the current task, retrospective estimates might have been especially precise because children had visual access to their completed construction, which served as immediate feedback. Moreover, the high accuracy of performance estimates may partly reflect the high level of task performance. In the Monkey Track task, performance was near ceiling across all cultural contexts. Since metacognitive bias is typically more pronounced in difficult tasks (Schneider, 1998), the simplicity of the current task might have reduced the likelihood of such bias. Future research should replicate these findings with a more challenging task.

Relations between metacognitive measurements

We explored the relationships between monitoring behaviors and task performance. Inspecting the model and comparing objects with the model were positively associated with task performance. Interestingly, the relationship between glancing and task performance was positive for low-performing children but negative for high-performing children. This pattern suggests that glancing may reflect a relatively basic (less thorough) form of monitoring that is informative for task performance but may be insufficient for achieving the highest levels of performance. For instance, a child may recognize from a distance (glancing) that a cup needs to be placed, but may need to inspect the model more closely to determine whether it is a small or a big cup. Overall, the coded monitoring behaviors contribute to children's task performance.

Moreover, lower metacognitive bias was weakly associated with longer monitoring duration but not with the number of monitoring occurrences. This pattern suggests that verbal metacognitive assessments (i.e., bias) and behavioral measures may capture partially distinct aspects of metacognition. This interpretation is consistent with recent studies showing that different components of metacognition are not necessarily correlated (Eberhart et al., 2025; Händel et al., 2025) and highlights the value of using multiple measures when comparing children's metacognition across cultural contexts to capture a broader range of metacognitive abilities.

Limitations and future research

Task performance was high in all cultural contexts (>80%), which may have limited the need to monitor behavior. However, children engaged in more than 8 monitoring behaviors per minute, which is similar to previous studies (Bryce & Whitebread, 2012; Buehler & Oeri, 2024; Eberhart et al., 2025). Examining how monitoring behavior changes with task difficulty would be an interesting direction for future research. Task difficulty could be manipulated by using models of varying complexity (Bryce & Whitebread, 2012), restricting task duration (Marulis & Nelson, 2021), or presenting children with an unsolvable task (Buehler & Oeri, 2026).

Task performance differed between cultural contexts. Chinese (95.37%) and U.S. children (89.20%) performed significantly better than Kenyan children (81.25%) on the Monkey Track task. As mentioned previously, to reliably compare metacognition between groups, task performance should ideally be identical, as metacognitive processes are less accurate in difficult tasks (Fleming & Lau, 2014). Future cross-cultural research may benefit from assessing metacognition in various domains, which may be less influenced by formal schooling and access to toys. For instance, motor tasks (e.g., Xia et al., 2022, 2023b) or perceptual tasks (e.g., Baer et al., 2021; Gonzales et al., 2022; van der Plas et al., 2022).

Socioeconomic status (SES) is a common confound in cross-cultural research. Indeed, SES was positively associated with metacognition among Western (French) children (Maximino-Pinheiro et al., 2024, 2026) and parents from high socioeconomic backgrounds tend to ask more metacognitive questions (Thompson & Foster, 2014). Thus, observed cross-cultural differences in metacognition may partly reflect socioeconomic disparities rather than cultural differences. Importantly, using available socioeconomic data from China and the U.S., we found no correlations among annual household income, monitoring occurrences, monitoring duration, metacognitive bias, or task performance (see Supporting Information Table S2). We were unable to collect annual household

income in Kenya because we had no direct access to the parents. While this suggests that our data were unlikely to be affected by socioeconomic status, future cross-cultural studies might assess and match samples on socioeconomic status (e.g., van der Plas et al., 2022).

Finally, we recognize that substantial variation can exist within cultural contexts, potentially exceeding differences between cultural contexts (The Culture Factor, n.d.). For instance, urban Chinese and urban U.S. children may be more similar to each other than urban and rural Chinese children. Thus, findings should be interpreted with caution when drawing conclusions about cultural values from country-level comparisons. Assessing cultural values directly in future research could provide a more fine-grained understanding of cultural variation than country-level comparisons alone.

Theoretical implications

The present study offers mixed support for the *cultural origins hypothesis* (Heyes et al., 2020). The *cultural origins hypothesis* maintains that cultural context predominantly affects explicit, rather than implicit, forms of metacognition. Explicit metacognition involves conscious processes that are held in working memory. Explicit metacognition is typically assessed using verbal measures (see Heyes et al., 2020). In comparison, metacognitive monitoring behaviors, our focus here, are a form of implicit metacognition and, as such, may be less conscious and require less working memory. For instance, glancing at the model may be a highly automated process that is not accompanied by a conscious representation, such as “*I am not sure which object comes next*”. Chinese children exhibited a different pattern of monitoring behaviors than Kenyan and U.S. children, raising the possibility that implicit monitoring behaviors are also affected by cultural learning.

However, while we find cross-cultural differences in metacognitive bias, they are likely attributable to differences in task performance. Thus, our findings do not replicate previous evidence for the cultural origins hypothesis based on explicit verbal assessments (Sunae Kim et al., 2021; Ordin et al., 2024; van der Plas et al., 2022; Xia et al., 2022, 2023b). At the same time, the observed cross-cultural differences in metacognitive behavior provide preliminary evidence that more implicit behavioral measures may be sensitive to culturally shaped learning experiences.

Taken together, the current study revealed both culturally-recurrent and culturally-variable aspects of children’s metacognitive behavior across three diverse cultural contexts. Consistent with prior research on U.S. and European samples, children in China and Kenya showed age-related increases in metacognitive behaviors between ages 6 and 10. At the same time, children from the three cultural contexts differed in their use of specific monitoring strategies. Chinese children, compared to both Kenyan and U.S. children, engaged more frequently in fine-grained, time-intensive monitoring behaviors such as inspecting the model and comparing individual objects. Future research should examine which culturally specific experiences contribute to these observed differences.

Data availability statement

Data is available on OSF (<https://osf.io/f8gw2/>).

Ethics approval statement

We obtained ethical approval from the institutional review boards of the University of California, Berkeley, the Institute of Psychology, Chinese Academy of Sciences, and the Kenyan National Commission for Science, Technology, and Innovation.

Funding statement

Florian Jonas Buehler was supported by the following grant: Swiss National Science Foundation, Grant/Award Number: P500PS 225245. Zhen Zhang was supported by CAS Key Laboratory of Behavioral Science, Institute of Psychology (E2CX1524YZ). Henriette Zeidler was supported by the European Union through a Marie-Sklodowska-Curie Postdoctoral Fellowship (101148929 — INCH).

CRediT authorship contribution statement

Florian Jonas Buehler: Writing – review & editing, Writing – original draft, Visualization, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Laura Simone Lewis:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Zhen Zhang:** Writing – review & editing, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization, Supervision, Resources. **Hanna Schleihau:** Writing – review & editing, Visualization, Validation, Methodology, Conceptualization. **Yufei Zong:** Writing – review & editing, Data curation, Investigation, Methodology. **Esther Herrmann:** Writing – review & editing, Methodology, Conceptualization. **Henriette Zeidler:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Data curation, Conceptualization, Methodology. **Jan Maximilian Engelmann:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to

influence the work reported in this paper.

Acknowledgments

We gratefully acknowledge the following people for their support in data collection and video coding: Yiran Du, Lina Huang, Rylie Harper, Sydney Tran, Stephen Kuom Elondo, John Lolo Edonga, Caroline Wanjiru Maina, and Dennis Wachira Maina.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jecp.2026.106691>.

References

- Arnett, J. J. (2008). The neglected 95%: Why american psychology needs to become less american. *American Psychologist*, 63(7), 602–614. <https://doi.org/10.1037/0003-066X.63.7.602>
- Baer, C., Ghetti, S., & Odic, D. (2021). Perceptual and memory Metacognition in Children. *Proceedings of the Annual Meeting of the Cognitive Science Perceptual and Memory Metacognition in Children*, 43(43), 3061–3067.
- Binah-Pollak, A. (2014). Discourses and practices of child-rearing in China: The bio-power of parenting in Beijing. *China Information*, 28(1), 27–45. <https://doi.org/10.1177/0920203X13517617>
- Brummelman, E., Nelemans, S. A., Thomaes, S., & Orobio de Castro, B. (2017). When parents' praise inflates, children's self-esteem deflates. *Child Development*, 88(6), 1799–1809. <https://doi.org/10.1111/CDEV.12936>
- Bryce, D., & Whitebread, D. (2012). The development of metacognitive skills: Evidence from observational analysis of young children's behavior during problem-solving. *Metacognition and Learning*, 7, 197–217. <https://doi.org/10.1007/s11409-012-9091-2>
- Buehler, F. J., & Oeri, N. (2024). Sneaky snake: Assessing metacognitive behavior in 5 to 6 year-olds with an unsolvable task. *Frontiers in Developmental Psychology*, 2. <https://doi.org/10.3389/fdps.2024.1454717>
- Buehler, F. J., & Oeri, N. (2026). Understanding temporal patterns of metacognitive monitoring and control in 5- to 6-year-olds: A latent class vector-autoregression analysis. *Developmental Psychology*, 62(4). <https://doi.org/10.1037/dev0002071>
- Buehler, F. J., Orth, U., Krauss, S., & Roebbers, C. M. (2025). Language abilities and metacognitive monitoring development: Divergent longitudinal pathways for native and non-native speaking children. *Learning and Instruction*, (95) <https://doi.org/10.1016/j.learninstruc.2024.102043>
- California Department of Education. (2025, January 10). *School Attendance Review Boards*. <https://www.cde.ca.gov/ls/ai/sb/>.
- Carr, M., Kurtz, B. E., Schneider, W., Turner, L. A., & Borkowski, J. G. (1989). Strategy acquisition and transfer among American and German children: Environmental influences on metacognitive development. *Developmental Psychology*, 25(5), 765–771. <https://doi.org/10.1037/0012-1649.25.5.765>
- Chao, R. K. (2000). The parenting of immigrant chinese and european american mothers: Relations between parenting styles, socialization goals, and parental practices. *Journal of Applied Developmental Psychology*, 21(2), 233–248. [https://doi.org/10.1016/S0193-3973\(99\)00037-4](https://doi.org/10.1016/S0193-3973(99)00037-4)
- Chen, J., & Park, A. (2021). School entry age and educational attainment in developing countries: Evidence from China's compulsory education law. *Journal of Comparative Economics*, 49(3), 715–732. <https://doi.org/10.1016/J.JCE.2020.12.004>
- Clegg, J. M., Wen, N. J., DeBaylo, P. H., Alcott, A., Keltner, E. C., & Legare, C. H. (2021). Teaching through collaboration: Flexibility and diversity in caregiver-child interaction across cultures. *Child Development*, 92(1), e56–e75. <https://doi.org/10.1111/CDEV.13443>
- Coffman, J. L., Ornstein, P. A., McCall, L. E., & Curran, P. J. (2008). Linking teachers' memory-relevant language and the development of children's memory skills. *Developmental Psychology*, 44(6), 1640–1654. <https://doi.org/10.1037/a0013859>
- Cottini, M., Basso, D., Pieri, A., & Palladino, P. (2021). Metacognitive monitoring and control in children's prospective memory. *Journal of Cognition and Development*, 22(4), 619–639. <https://doi.org/10.1080/15248372.2021.1916500;CTYPE=STRING:JOURNAL>
- Coughlin, C., Hembacher, E., Lyons, K. E., & Ghetti, S. (2015). Introspection on uncertainty and judicious help-seeking during the preschool years. *Developmental Science*, 18(6), 957–971. <https://doi.org/10.1111/desc.12271>
- Destan, N., Hembacher, E., Ghetti, S., & Roebbers, C. M. (2014). Early metacognitive abilities: The interplay of monitoring and control processes in 5- to 7-year-old children. *Journal of Experimental Child Psychology*, 126, 213–228. <https://doi.org/10.1016/j.jecp.2014.04.001>
- Dinno, A. (2024). *dunn.test: Dunn's test of multiple comparisons using rank sums* (Version 1.3.6) [R package]. CRAN. <https://CRAN.R-project.org/package=dunn.test>.
- Dunlosky, J., Mueller, M. L., & Thiede, K. W. (2016). Methodology for investigating human metamemory: problems and pitfalls. In J. Dunlosky, & S. K. Tauber (Eds.), *The Oxford Handbook of Metamemory*. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199336746.013.14>
- Eberhart, J., Murayama, K., Sakaki, M., & Bryce, D. (2025). Metacognitive monitoring in early elementary school-aged children: Task dependency in monitoring judgments, task consistency in monitoring behaviours. *Cognitive Development*, 74, Article 101561. <https://doi.org/10.1016/J.COGEDEV.2025.101561>
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8(443), 1–9. <https://doi.org/10.3389/fnhum.2014.00443>
- Gardier, M., & Geurten, M. (2025). Is uncertainty in the eyes or in parents' talk? linking an eye-tracking measure of toddlers' core metacognition to parental metacognitive talk. *Child Development*, 96(4), 1385–1394. <https://doi.org/10.1111/cdev.14237>
- Geurten, M., & Léonard, C. (2023). Relations between parental metacognitive talk and children's early metacognition and memory. *Journal of Experimental Child Psychology*, 226. <https://doi.org/10.1016/J.JECP.2022.105577>
- Gonzales, C. R., Mercurieff, A., McClelland, M. M., & Ghetti, S. (2022). The development of uncertainty monitoring during kindergarten: Change and longitudinal relations with executive function and vocabulary in children from low-income backgrounds. *Child Development*, 93(2), 1–16. <https://doi.org/10.1111/cdev.13714>
- Grammer, J., Coffman, J. L., & Ornstein, P. (2013). The effect of teachers' memory-relevant language on children's strategy use and knowledge. *Child Development*, 84(6), 1989–2002. <https://doi.org/10.1111/CDEV.12100>
- Gunderson, E. A., Gripshover, S. J., Romero, C., Dweck, C. S., Goldin-Meadow, S., & Levine, S. C. (2013). Parent praise to 1- to 3-year-olds predicts children's motivational frameworks 5 years later. *Child Development*, 84(5), 1526–1541. <https://doi.org/10.1111/CDEV.12064>
- Händel, M., Nett, U., Bryce, D., & Dresel, M. (2025). Unraveling the reciprocal effects and micro-development of metacognitive monitoring strategies and monitoring judgments: A micro-genetic diary study. *Learning and Individual Differences*, 122. <https://doi.org/10.1016/j.lindif.2025.102748>
- Hembacher, E., & Ghetti, S. (2014). Don't look at my answer: Subjective uncertainty underlies preschoolers' exclusion of their least accurate memories. *Psychological Science*, 25(9), 1768–1776. <https://doi.org/10.1177/0956797614542273>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), 61–83. <https://doi.org/10.1017/S0140525X0999152X>
- Heyes, C., Bang, D., Shea, N., Frith, C. D., & Fleming, S. M. (2020). Knowing ourselves together: The cultural origins of metacognition. *Trends in Cognitive Sciences*, 24(5), 349–362. <https://doi.org/10.1016/j.tics.2020.02.007>
- Hou, S., Wang, Y., Cai, P., Chen, R., & Wang, R. (2021). What parents bring to preschool children? Parental instructive speech and gestures, children's learning and cultural differences. *Children and Youth Services Review*, 127. <https://doi.org/10.1016/j.childyouth.2021.106078>

- Huntsinger, C. S., & Jose, P. E. (2009). Parental involvement in children's schooling: Different meanings in different cultures. *Early Childhood Research Quarterly*, 24(4), 398–410. <https://doi.org/10.1016/j.ecresq.2009.07.006>
- Keller, H. (2012). Autonomy and relatedness revisited: Cultural manifestations of universal human needs. *Child Development Perspectives*, 6(1), 12–18. <https://doi.org/10.1111/j.1750-8606.2011.00208.x>
- Kenya National Bureau of Statistics. (2019). *Kenya Population and Housing Census*. <https://www.knbs.or.ke/reports/kenya-census-2019/>.
- Kim, S. (2015). ppcor: Partial and semi-partial (part) correlation (Version 1.1) [R package]. CRAN. <https://CRAN.R-project.org/package=ppcor>.
- Kim, S., Senju, A., Sodian, B., Paulus, M., Itakura, S., Okuno, A., Ueno, M., & Proust, J. (2021). Memory monitoring and control in Japanese and German preschoolers. *Memory and Cognition*. <https://doi.org/10.3758/s13421-021-01263-1>
- Kim, Y. H., Chiu, C. Y., Peng, S., Cai, H., & Tov, W. (2010). Explaining east-west differences in the likelihood of making favorable self-evaluations: The role of evaluation apprehension and directness of expression. *Journal of Cross-Cultural Psychology*, 41(1), 62–75. <https://doi.org/10.1177/0022022109348921>
- Kinoti, H. W. (2010). African ethics: Gikuyu traditional morality. In *African ethics*. Brill. <https://philpapers.org/rec/KINAEG>.
- Langenhoff, A. F., Thompson, B. D., Srinivasan, M., & Engelmann, J. M. (2025). Disagreement drives metacognitive development. *Trends in Cognitive Sciences*. <https://doi.org/10.1016/j.tics.2025.05.014>
- Lenth, R. V. (2025). emmeans: Estimated marginal means, aka least-squares means (Version 1.10.7) [R package]. CRAN. <https://CRAN.R-project.org/package=emmeans>.
- Lipko, A. R., Dunlosky, J., & Merriman, W. E. (2009). Persistent overconfidence despite practice: The role of task experience in preschoolers' recall predictions. *Journal of Experimental Child Psychology*, 103(2), 152–166. <https://doi.org/10.1016/j.jecp.2008.10.002>
- Lyons, K. E., & Ghetti, S. (2011). The development of uncertainty monitoring in early childhood. *Child Development*, 82(6), 1778–1787. <https://doi.org/10.1111/j.1467-8624.2011.01649.x>
- Marulis, L. M., & Nelson, L. J. (2021). Metacognitive processes and associations to executive function and motivation during a problem-solving task in 3–5 year olds. *Metacognition and Learning*, 16(1), 207–231. <https://doi.org/10.1007/s11409-020-09244-6>
- Max Planck Institute for Psycholinguistics The Language Archive. (2024). ELAN. <https://archive.mpi.nl/tla/elan>.
- Maximino-Pinheiro, M., Menu, I., Boissin, E., Brunet, L. A., Barone, C., & Borst, G. (2024). Metacognition as a mediator of the relation between family SES and language and mathematical abilities in preschoolers. *Scientific Reports*, 14. <https://doi.org/10.1038/s41598-024-60972-0>
- Maximino-Pinheiro, M., Rémeau, M., Le Stanc, L., Barone, C., & Borst, G. (2026). Reducing educational inequalities through metacognition? evidence from a classroom intervention in kindergarten. *Metacognition and Learning*, 21(3). <https://doi.org/10.1007/s11409-025-09452-y>
- McCulloch, C. E., & Searle, S. R. (2000). *Generalized, linear, and mixed models*. John Wiley & Sons.
- Merkin, R., Taras, V., & Steel, P. (2014). State of the art themes in cross-cultural communication research: A systematic and meta-analytic review. *International Journal of Intercultural Relations*, 38(1), 1–23. <https://doi.org/10.1016/J.IJINTREL.2013.10.004>
- Middleton, J., & Kershaw, G. (2017). *The Kikuyu and Kamba of Kenya: East central Africa Part V*. Routledge, 10.4324/9781315313139/KIKUYU-KAMBA-KENYA-JOHN-MIDDLETON-GREET-KERSHAW.
- Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *Psychology of Learning and Motivation - Advances in Research and Theory*, 26, 125–173.
- Ng, F.-F.-Y., Pomerantz, E. M., & Lam, S. (2007). European American and Chinese parents' responses to children's success and failure: implications for Children's responses. *Developmental Psychology*, 43(5), 1239–1255. <https://doi.org/10.1037/0012-1649.43.5.1239>
- Ng, F.-F.-Y., & Wei, J. (2020). Delving into the minds of Chinese parents: What beliefs motivate their learning-related practices? *Child Development Perspectives*, 14(1), 61–67. <https://doi.org/10.1111/CDEP.12358>
- Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, 162, 31–38. <https://doi.org/10.1016/j.jecp.2017.04.017>
- Ordin, M., El-Dakhs, D. A. S., Tao, M., Chu, F., & Polyanskaya, L. (2024). Cultural influence on metacognition: Comparison across three societies. *Humanities and Social Sciences Communications*, 11(1), 1492. <https://doi.org/10.1057/s41599-024-04013-1>
- Price, N. (1996). The changing value of children among the Kikuyu of Central Province, Kenya. *Africa*, 66(3). <https://doi.org/10.2307/1160960>
- R Core Team. (2025). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Ran, A. (2001). Travelling on parallel tracks: Chinese parents and English teachers. *Educational Research*, 43(3), 311–328. <https://doi.org/10.1080/00131880110081062>
- Roebbers, C. M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, 45, 31–51. <https://doi.org/10.1016/j.dr.2017.04.001>
- Schneider, W. (1998). Performance prediction in young children: Effects of skill, metacognition and wishful thinking. *Developmental Science*, 1(2), 291–297. <https://doi.org/10.1111/1467-7687.00044>
- Schraw, G. (2009). A conceptual analysis of five measures of metacognitive monitoring. *Metacognition and Learning*, 4(1), 33–45. <https://doi.org/10.1007/s11409-008-9031-3>
- Sedikides, C., Gaertner, L., & Cai, H. (2015). On the panculturality of self-enhancement and self-protection motivation: The case for the universality of self-esteem. *Advances in Motivation Science*, 2, 185–241. <https://doi.org/10.1016/bs.adms.2015.04.002>
- Shanley, J. R., Mutiso, V., Musyimi, C., Armistead, L., Olumbe, R., Ishiekwe, M. N., & Ndeti, D. (2024). Kenyans' perspectives on parenting roles and strategies used to raise young children in Kenya. *Journal of Family Psychology*, 38(7), 1007–1016. <https://doi.org/10.1037/fam0001259>
- Stelz, S., & Altgassen, M. (2024). Prospective and retrospective metacognitive judgments of prospective memory performance across the lifespan. *Journal of Cognition and Development*, 1–22. <https://doi.org/10.1080/15248372.2024.2441689>
- Sun, J., & Ryder, A. G. (2016). The Chinese experience of rapid modernization: Sociocultural changes, psychological consequences? *Frontiers in Psychology*, 7(APR), Article 109319. <https://doi.org/10.3389/fpsyg.2016.00477/BIBTEX>
- Thalmayer, A. G., Toscanelli, C., & Arnett, J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist*, 76(1), 116–129.
- The Culture Factor. (n.d.). *Country comparison tool*. Retrieved February 25, 2025, from <https://www.theculturefactor.com/country-comparison-tool?countries=china%2Ckenya%2Cunited+states>.
- Thompson, R. B., & Foster, B. J. (2014). Socioeconomic status and parent-child relationships predict metacognitive questions to preschoolers. *Journal of Psycholinguistic Research*, 43(4), 315–333. <https://doi.org/10.1007/s10936-013-9256-4>
- United States Census Bureau. (2023). *American Community Survey 1-Year Estimates*. <https://data.census.gov/profile/>.
- van der Plas, E., Zhang, S., Dong, K., Bang, D., Li, J., Wright, N. D., & Fleming, S. M. (2022). Identifying cultural differences in metacognition. *Journal of Experimental Psychology: General*, 151(12), 3268–3280. <https://doi.org/10.1037/xge0001209>
- Voronov, M., & Singer, J. A. (2002). The myth of individualism-collectivism: A critical review. *The Journal of Social Psychology*, 142(4), 461–480. <https://doi.org/10.1080/00224540209603912>
- Wang, Y. Z., Wiley, A. R., & Chiu, C. Y. (2008). Independence-supportive praise versus interdependence-promoting praise. *International Journal of Behavioral Development*, 32(1), 13–20. <https://doi.org/10.1177/0165025407084047>
- Whiting, B. B. (1996). The effect of social change on concepts of the good child and good mothering: A study of families in Kenya. *Ethos*, 24(1), 3–35.
- Wong, Y. J., Wang, S.-Y., Klann, E. M., & Wong, J. (2018). The emperor with no clothes: A critique of collectivism and individualism archives of scientific psychology. *Archives of Scientific Psychology*, 6, 251–260. <https://doi.org/10.1037/arc0000059>
- Wu, P., Robinson, C. C., Yang, C., Hart, C. H., Olsen, S. F., Porter, C. L., Jin, S., Wo, J., & Wu, X. (2002). Similarities and differences in mothers' parenting of preschoolers in China and the United States. *International Journal of Behavioral Development*, 26(6), 481–491. <https://doi.org/10.1080/01650250143000436>

- Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2023a). Children's overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development*, 1–22. <https://doi.org/10.1111/cdev.14042>
- Xia, M., Poorthuis, A. M. G., & Thomaes, S. (2023b). Why do young children overestimate their task performance? a cross-cultural experiment. *Journal of Experimental Child Psychology*, 226. <https://doi.org/10.1016/j.jecp.2022.105551>
- Xia, M., Poorthuis, A. M. G., Zhou, Q., & Thomaes, S. (2022). Young children's overestimation of performance: A cross-cultural comparison. *Child Development*, 93(2), e207–e221. <https://doi.org/10.1111/cdev.13709>
- Xu, Y., & Hamamura, T. (2014). Folk beliefs of cultural changes in China. *Frontiers in Psychology*, 5(SEP), Article 104369. <https://doi.org/10.3389/fpsyg.2014.01066/BIBTEX>
- Zhu, N., & Chang, L. (2019). Education and parenting in China. In E. Sorbring, & J. E. Lansford (Eds.), *School Systems, Parent Behavior, and Academic Achievement: An international perspective* (pp. 15–28). Springer. https://doi.org/10.1007/978-3-030-28277-6_2.