Studies in Educational Evaluation 81 (2024) 101339

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

Self-regulated learning: Validating a task-specific questionnaire for children in elementary school

R. Ebbes [a,d,*], J.A. Schuitema [a,d], H.M.Y. Koomen [a,d], B.R.J. Jansen [b,d], M. Zee [a,c,d]

[a] Research Institute of Child Development and Education, University of Amsterdam, the Netherlands
[b] Developmental Psychology, University of Amsterdam, the Netherlands
[c] Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, the Netherlands
[d] Research Priority Area Yield, University of Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Self-regulated learning
Math task
Elementary education

ABSTRACT

This paper describes the development and initial validation of the Cognition and Emotion/Motivation Regulation (CEMOR) questionnaire, a task-specific questionnaire for upper elementary school students that measures self-regulated learning (SRL). Using a multistep procedure, 22 items were developed, divided over five theory-informed dimensions (Planning, Monitoring, Cognition Control, Emotion/Motivation Control, and Reflecting). The CEMOR was applied in a math context. Children from grades 3–6 ($N = 547$, 54.7 % females) completed the CEMOR. Confirmatory factor analyses indicated that the five proposed scales have adequate to good model fit, with factor loadings ranging from .54 to .83, and acceptable to good composite reliability (ρ range = .75–.85). To find further validity support, the SRL scales were correlated with students' performance on a math task, experienced emotions, and level of motivation during the task. Most correlations were statistically significant and in the expected direction. Hence, the CEMOR questionnaire shows promise as a new SRL instrument for elementary education.

Self-regulated learning (SRL) is an active, constructive process in which learners set goals for their learning and then attempt to monitor, regulate, and control their cognition, emotion, motivation, and behavior, guided and constrained by their goals and contextual features (Pintrich, 2000). Children's SRL is positively related to their academic success (Dent & Koenka, 2016), considered an essential skill for lifelong learning (Cornford, 2002), and often mentioned as an important educational goal (Bolhuis, 2003). Already in elementary education, children can and do engage in accurate SRL activities (Perry et al., 2002; Schneider, 2008). For instance, SRL plays an important role in the mathematical domain when children select strategies or monitor the accuracy of their answers (De Corte et al., 2011). It has even been argued that the elementary school context is one of the most important to learn SRL (Vandevelde et al., 2013). During this period, children have not yet fully developed the studying habits and beliefs about their own abilities (Usher & Pajares, 2008) that might hinder future SRL (Dignath & Büttner, 2008). Moreover, SRL may be easier to learn in elementary education because it allows for more gradual practice in a less complex and demanding context than secondary education (Meneghetti et al., 2007).

This emphasizes the need for SRL-research with an explicit focus on elementary education. Specifically, research is warranted on when and how different SRL-processes develop during childhood. This kind of research can inform research and practice on what kind of SRL proficiency is feasible in elementary education. Moreover, it helps to understand the factors that relate to SRL development, such as specific cognitive processes, social interactions, and instructional strategies. This knowledge can guide development of evidence-based instructional practice, curricula that foster self-regulation, and interventions that target specific SRL processes. A necessity for this kind of research is a reliable and valid method to assess children's SRL.

Currently, there is a variety of SRL instruments available to researchers (for an overview see Koivuniemi et al., 2021). However, very little of these instruments cover the full concept of SRL, with motivational and emotional learning aspects particularly lacking (Koivuniemi et al., 2021). Moreover, to our knowledge, available instruments are either relatively time-consuming and hard to administer on a large scale (e.g., think-aloud protocols or observation instruments), or they

* Correspondence to: University of Amsterdam, Research Institute of Child Development and Education, Nieuwe Achtergracht 129b, 1018 WS Amsterdam, the Netherlands.
E-mail address: r.ebbes@uva.nl (R. Ebbes).

https://doi.org/10.1016/j.stueduc.2024.101339
Received 13 October 2022; Received in revised form 17 January 2024; Accepted 24 January 2024
Available online 17 February 2024

measure SRL as a general, relatively stable ability. The view of SRL as a more fluid ability, that can differ between or even within specific situations (Boekaerts & Corno, 2005), underlies our decision to develop a task-specific questionnaire to evaluate SRL. To our knowledge, such a questionnaire does not yet exist. Therefore, the aim of the current study is to develop and initially validate a task-specific questionnaire for upper elementary education that measures SRL. This questionnaire will be applied to a math task here but is designed to be applied in other academic domains as well.

1. Defining self-regulated learning

Due to a long research tradition in various fields of psychology (see e.g., Winne, 2017 for an overview), there are a number of different models of SRL that propose different constructs and different conceptualizations of SRL (see e.g., Panadero, 2017; Puustinen & Pulkkinen, 2001 for an overview). Several of those models make a distinction between phases of regulation (e.g., Pintrich, 2000; Zimmerman, 2000) and areas of regulation (Pintrich, 2000). The distinction between phases and areas of SRL made in the model by Pintrich (2000) was used as the foundation of our instrument.

Pintrich (2000) makes a distinction between the phases of planning, monitoring, control, and reflecting. The first phase, planning, consists of processes such as task orientation, goal setting, and strategic planning. During task orientation, children analyze the task at hand and try to get a grasp on the task demands and objectives (Desoete, 2008; Vandevelde et al., 2013). Moreover, children use metacognitive knowledge of the task and memories about past performance on similar tasks to judge the difficulty of the task, sometimes referred to as ease of learning judgments (Pintrich, 2000). Based on such judgments of competence and children's perception of the value and interest of the task, children set their own goals and decide, for instance, on their level of engagement (Pintrich, 2000). The last aspect of the planning phase entails strategic planning, where children select strategies that are most appropriate for the task and the setting (Zimmerman, 2000). Although the use of appropriate strategies is necessary to successfully complete a task, it is the decision to use them, stop using them, or switch between strategies that is seen as regulation (Pintrich, 2000).

The second phase concerns various monitoring processes. Monitoring involves the ongoing assessment of the quality of task performance and the degree to which performance is progressing toward a desired goal, thereby informing the learner whether modification of currently applied strategies is needed (Pintrich, 2000; Vandevelde et al., 2013). Children can for instance monitor their degree of comprehension, their feeling of knowing or their learning (Moos & Azevedo, 2008; Pintrich, 2000; Vandevelde et al., 2013).

The third phase entails the efforts to control and regulate different aspects of the self and the task. Based on monitoring, children can deduce that their current approach is not working, and consequently decide to change strategies or to ask for help. Last, the fourth phase, reflecting, involves children's judgments and evaluation of their performance on the task as well as their attributions of performance (Pintrich, 2000). Children can for instance check the accuracy and the completeness of their performance (Vandevelde et al., 2013), reflect on the effectiveness of their used strategies (Desoete, 2008), and attribute causal factors to their level of success (Pintrich, 2000; Zimmerman, 2000).

Independent of the phases of regulation, the model by Pintrich (2000) makes a distinction between areas that children can attempt to regulate during their learning. The first area that Pintrich describes is that of cognition. Children proficient in cognitive SRL have a large arsenal of cognitive strategies that they can readily and skillfully deploy to accomplish different academic tasks (Wolters, 2003). Moreover, they have the ability to select, monitor, and regulate their use of those strategies when engaged in academic tasks (Wolters, 2003).

Pintrich (2000) describes motivation and emotions as a second area subject to regulation. Motivation entails the willingness to engage in and persist at a task, and is influenced, for example, by children's level of interest, feelings of self-efficacy, or their drive to master a certain activity (Wolters, 2003). Consequently, the regulation of motivation entails the monitoring and regulation of both motivation and the processes responsible for motivation (Wolters, 2003). For example, children can try to increase their feelings of self-efficacy and consequently their level of motivation through self-talk (e.g., telling themselves "You can do it!"). Closely linked to motivation are children's emotions. Emotions can impact children's motivation, learning, and performance (Pekrun et al., 2011). If a task is perceived as difficult, ambiguous, or complex, it can trigger negative emotions such as frustration or anxiety. In this case, regulation activities are necessary to prevent or limit the negative impact this might have on learning and performance (Boekaerts, 2011). For instance, when children feel frustrated or angry, they can try to calm themselves by focusing on their breathing. Motivation and emotions are considered as closely related and interconnected dimensions within the model proposed by Pintrich. Although they are distinct constructs, they are often discussed together due to their intertwined nature and their shared influence on learners' engagement, behaviors, and overall learning outcomes. By combining them into a single area, the model recognizes their collective impact on self-regulation processes.

Other areas that Pintrich describes as suitable for SRL are the available resources and the context. When regulating resources, children try, for instance, to monitor and control the available time for a task and the amount of help that can be called upon to complete a learning task. Children regulating the learning context can, for instance, try to make changes to their working space or the facilities they use to learn. In most Dutch elementary schools, teachers regulate resources and the learning context, thereby leaving little room for children to self-regulate these areas. We therefore leave the areas recourses and context outside the scope of our instrument.

2. Measuring self-regulated learning

Several methods can be used to measure SRL (Boekaerts & Corno, 2005). Often, a distinction is made between off-line and on-line methods to measure SRL (Schellings, 2011). *Offline* methods (e.g., self-report questionnaires or teacher ratings) assess SRL before or after a task performance. Benefits of offline instruments are that they are often easy to administer, take little time to administer/complete (Cromley & Azevedo, 2011), and do not disturb learners during their learning activities (Van Hout-Wolters & Schellings, 2009). A disadvantage of most offline instruments is that they measure SRL as a general trait and consequently ask children to generalize about SRL across multiple times and situations (Endedijk et al., 2016). This makes it unclear which situations children recall when completing such a questionnaire and how they weigh different experiences (Dinsmore et al., 2008). Moreover, this recollection can be prone to memory failure and distortions (Nisbett & Wilson, 1977; Veenman, 2011a).

Less prone to memory failure are *online* instruments (e.g., think-aloud protocols, observations, or trace methods), that measure SRL during task performance (Endedijk et al., 2016). However, these instruments also have their limitations. For instance, think-aloud protocols, where children are asked to verbalize their cognitive processes while performing a task, may still be incomplete as some processes are unconscious (Bainbridge, 1979) or difficult to verbalize (Schuck & Leahy, 1966). Moreover, verbalizing during a task asks for additional cognitive processing which might impact the original performance, especially in complex tasks with a high cognitive load (Fox et al., 2011). Alternative on-line methods at a more micro-level, such as observations or trace methodologies (where traces of student work, such as highlighted sections in text or comments written in the margin, are collected and studied), often only capture easily observed activities or overt behavior, and not the thoughts and motives underlying children's behavior (Veenman, 2011a). In addition, on-line methods are often time and labor-intensive, making them less suitable for use in larger samples (Vandevelde et al., 2013).

Another limitation of currently available instruments is that they often measure only a rather small subset of SRL practices. For instance, the Patterns of Adaptive Learning Scales (PALS; Midgley et al., 2000) assesses students' goal orientation and the Junior Metacognitive Awareness Inventory (Jr. MAI; Sperling et al., 2002) measures children's metacognitive skills. These questionnaires are appropriate for assessing their respective constructs. However, they are insufficient for researchers and practitioners interested in a broader picture of SRL including the regulation of emotions and motivation. More comprehensive is the Children's Perceived use of Self-Regulated Learning Inventory (CP-SRLI; Vandevelde et al., 2013), that measures both cognitive and emotional regulation during academic homework (i.e., in the home context). However, to allow for assessment of SRL during a normal school day, an instrument is necessary that measures SRL in the school context. To our knowledge, such an instrument does not yet exist.

Self-regulated learning can be seen as domain-specific (Boekaerts & Corno, 2005). Distinct domains, for instance history and mathematics, differ with respect to the nature and structure of their tasks and subject matter, which influences how and how easily children can regulate these processes (Poitras & Lajoie, 2013). Even within domains, task characteristics can facilitate or constrain SRL (Pintrich, 2000). For instance, children's knowledge about when and how to use various strategies is often situation-specific (Veenman, 2011b) and children's motivation to invest in SRL is to a large degree dependent on whether they see a specific task as useful and interesting (Pintrich, 2000). In short, SRL can be seen as a fluent ability that fluctuates between and sometimes even within specific situations (Boekaerts & Corno, 2005).

Considering the perspective of SRL as a fluent ability that is situation specific, we developed an offline questionnaire that measures children's SRL during a specific task. This makes this instrument more suitable for children, as children often have trouble to generalize about their SRL ability over multiple contexts. By administering the questionnaire immediately after children have completed this task, we minimize potential memory issues associated with questionnaires (Veenman, 2011a), while maintaining their benefits for large scale assessment. Moreover, the task-specific nature of this new instrument allows for a more fine-grained analysis of SRL in specific situations, which is necessary to better understand the situation specific nature of SRL (Alexander et al., 2015). The items of this new instrument were developed to be generic, with the phrasing of our items inferring about children's self-regulation while making 'the exercises'. This allows for the instrument to be used within various domains and with different tasks. For the validation of this instrument, we start with a single domain and a single task.

We chose for a math task. Most math tasks require the use of multiple mathematical competencies, for instance problem solving, modelling, and mathematical reasoning (Niss & Højgaard, 2019), making math a very suitable domain for SRL (De Corte et al., 2011). We focused on the task-specific context of word-problems, as these often ask for a complex solution process (Jonassen, 2003) that requires students to integrate several cognitive processes (Jitendra et al., 2007), which necessitates a high degree of cognitive SRL. Additionally, word-problems are often experienced as difficult, sometimes resulting in feelings of anxiety (Ashcraft, 2002) or frustration (Wigfield & Meece, 1988), which gives a good opportunity to study the emotional regulation students might employ. We selected challenging tasks, increasing the probability that also those proficient in math word problems need to apply SRL.

3. SRL and academic performance, motivation, and emotions

In general, psychometric quality of new instruments is investigated through factor and reliability analyses. Additional information about the validity of the new instrument can be gathered by investigating the relationship between measured SRL ability and constructs that are hypothesized to be related to SRL. To this end, we also investigated the relationship of SRL with task performance, motivation, and emotions.

SRL is associated with better academic performance, with high-achieving students using SRL strategies more frequently and effectively than their lower achieving peers (Dent & Koenka, 2016). There are several ways the process of SRL could benefit children's academic performance. For instance, children orienting themselves on a task are more likely to pay attention to and consequently meet demands and objectives of the task. Children monitoring the quality of their task performance are more likely to spot and thereby correct possible mistakes. Moreover, children reflecting on their performance are more likely to learn from their mistakes, thereby potentially improving their future performance. For these reasons, we expect a positive relationship between the SRL scales and children's performance on the math task.

Our conceptualization of SRL also encompasses the monitoring and regulation of emotions and motivation. In this case, successful SRL is likely to effectuate an increase in motivation (Pekrun et al., 2011). Similarly, high motivation makes it more likely that children will self-regulate their learning in order to do well (Pintrich, 2000). The relationship between SRL and motivation is therefore hypothesized to be positive.

In contrast, successful SRL is likely to result in a decrease in negative emotions (Pekrun et al., 2011). Similarly, negative emotions like boredom and hopelessness deactivate children, interfere with task focus (Tyson et al., 2009), and can therefore be argued to decrease the amount and quality of children's SRL. The relationship between SRL and deactivating negative emotions like boredom and hopelessness is therefore hypothesized to be negative.

Compared to this, the relationship between SRL and negative activating emotions like anxiety and anger is less clear cut. For instance, anger and anxiety can reduce cognitive resources available for task performance (Pekrun et al., 2007), which can be argued to negatively impact children's SRL. However, anxiety can also strengthen motivation to invest effort and avoid failure (Pekrun et al., 2007), which makes it more likely that children will try to minimize the risk of making errors by self-regulating their learning. Anger can be a powerful motivator to overcome current obstacles, to persist, and keep trying (Pekrun et al., 2002), which necessitates SRL. Moreover, making mistakes can result in feelings of anger and anxiety and (unrelated to these feelings) cue children to improve their process through SRL. In short, the relationship between SRL and activating negative emotions like anxiety and anger can be both positive and negative.

4. Present study

In the current study, we describe the development of the Cognition and Emotion/Motivation Regulation (CEMOR) questionnaire, a relatively short instrument that can be used to measure the cognition and emotion/motivation regulation of children in elementary education. To capture the situation-specific nature of SRL, the CEMOR questionnaire is task-specific, measuring the level of SRL children applied during a specific task, in this case a math task To assess the validity and reliability of this instrument, we investigated its factor structure and internal consistency. Moreover, we explored the relationships between CEMOR-scores and several variables hypothesized to be related to SRL ability (i.e., task performance, negative emotions, and motivation). We expect the SRL scales to have a positive relationship with children's performance on the math task and motivation, and a negative relationship with boredom and hopelessness. The relationship with anger and anxiety can be theorized to be both negative or positive.

5. Method

5.1. Participants

Our sample consisted of children in Dutch upper elementary education (grades 3–6). This sample was recruited by reaching out via email or phone to teachers currently teaching an upper elementary school class. We contacted teachers who participated in earlier research projects and teachers within the personal network of the researchers. This group was supplemented with teachers invited through cold calling a random sample of schools throughout the Netherlands. Interested teachers were asked to forward an information letter and consent form to the parents of all children in their classrooms. Of all teachers that were contacted, 40 (32 %) agreed to reach out to the parents of the children in their class. Of these children, 55 % received consent from their parents to participate and consequently took part in this project.

In total, 547 children ($M_{age} = 10.61$, $SD = 0.97$, range 8–13 years; 54.7 % females) from 18 schools and 40 classes (range 1–4 classes per school) participated. Participants were enrolled in grades 3 ($n = 57$), 4 ($n = 181$), 5 ($n = 238$) and 6 ($n = 71$), respectively. Most children were born in the Netherlands (96.2 %) and raised by Dutch-speaking caregivers (92.9 %). Children were asked what ethnic background they most identified with, which resulted in the following distribution: 92.1 % Dutch, 1.5 % Moroccan, 1.3 % Turkish, 0.7 % Surinamese, and 4.4 % other ethnicities.

5.2. Instruments

5.2.1. Math task

Prior to answering questions about their SRL, children were asked to work on a math task. Exercises were used from the Dutch version of the 2014 W4Kangoeroe math competition (w4kangoeroe.nl) for grades 3/4 (WizKID) and grades 5/6 (WizSMART). This annual math competition consists of a set of 24 word-problems that start easily, but gradually become more difficult. An example of a WizKID item is: “Piggy Oink loves melons and carrots. Every day he eats either 9 carrots, or 2 melons, or 1 melon and 4 carrots. In one week, Oink ate 30 carrots. How many melons did he eat that week?”. An example of a WizSMART item is: “Grandma Emma has six grandchildren. All grandchildren have a different age. If you add the ages of all the grandchildren, the result is 120. Sara is the oldest. What is the minimum age of Sara?”. For each item, children had to choose from five multiple choice options. The number of correct answers was used as indicator for the performance on this math task.

5.2.2. Demand scale

We asked children to indicate how hard they thought the math exercises were via a Demand scale. For this scale we developed five items (α current study = .82) that were answered on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). An example item is: “I thought these exercises were hard”.

5.2.3. CEMOR-questionnaire

A multistep procedure was used to develop the CEMOR questionnaire. First, an initial item pool was developed based on current definitions and operationalizations of SRL in the literature. Second, this item pool was reviewed by three educational researchers and three in-service elementary school teachers. Third, items were inspected using cognitive interviews. Last, items were pilot tested in a small-scale sample of 162 children.

Scale Development. We developed scales for each of the SRL phases (planning, monitoring, control, and reflecting). Regarding the different areas that can be regulated, we chose to include the regulation of cognition and the regulation of emotion/motivation. The Pintrich model also includes regulation of available resources and context but in most Dutch elementary schools, teachers regulate resources and the learning context, thereby leaving little room for children to self-regulate these areas. To design a parsimonious questionnaire, the areas were merged in one scale for each phase except for the control phase. For the control phase, separate control and emotion/motivation scales were included because the related strategies are too distinct to merge. For the scale Emotion/motivation Control, items were based on items from the AVSI questionnaire (McCann & Turner, 2004), an instrument that measures how participants enhance their self-efficacy and reduce their stress. For the scales Planning, Monitoring, Cognition Control, and Reflecting items were constructed based on current definitions and operationalizations in the literature. As the current instrument specifically targets children in upper elementary education, special attention was given to ensure that wording and phrasing were appropriate for this age-group. Each question in the questionnaire specifically referred to the task that was made prior to the questionnaire. For instance, most items started with the words “While making these exercises, I …”. All items were developed as statements that could be answered on a 5-point Likert scale ranging from 1 (totally disagree) to 5 (totally agree). This development process resulted in a first item pool of 57 items.

Item Review. Six reviewers (three educational scientists and three elementary school teachers) were asked to assess the items on content alignment with the construct, suitability for children in grades 3–6, and clarity of wording. Based on their feedback, several items were rephrased to better suit the vocabulary of children in grades 3–6 and to improve clarity of wording. Last, items were rephrased so that items from every scale started with the same words. This review of items resulted in an item pool of 40 items.

Cognitive Interviews. Cognitive interviews provide rich data about whether children’s interpretations of self-report items are consistent with researchers’ assumptions (Woolley et al., 2006). Moreover, such interviews provide insight into whether the items do not exceed participants’ cognitive ability to read, interpret, and respond to the items (Karabenick et al., 2007). Five randomly selected children from two classes (grade 4 and 5, three females, two males, $M_{\text{age}} = 10.2$, $SD = 0.67$), not included in the pilot or large-scale assessment, performed the math tasks and consequently responded to the CEMOR items. Participants were asked to (a) read the question aloud; (b) explain or paraphrase the question; (c) read the answer options and choose an answer; and (d) explain why they chose that answer. For most of the items, children were able to interpret the items correctly. Misinterpreted items were slightly revised regarding wording and phrasing to better facilitate understanding. For instance, children responded to the statement “I asked myself if I was understanding the material” by confirming they understood the exercises. As this was not our intended interpretation, we changed this item to “While making these exercises, I checked whether I understood what I was doing”. During this round of interviews, no items were discarded.

Pilot Testing. Last, the math tasks and CEMOR instrument were pilot tested in a sample of 162 children ($M_{age} = 10.93$, $SD = 1.01$, range 9–13 years; 54.9 % females) from four 4th grade, two 5th grade, and two 6th grade classes. The math task difficulty was found to be of the right level. Six CEMOR items showed no inter-item correlations above .40 with any other item and were discarded. CFA’s modification indices showed that the Cognitive Control items were also good indicators of the Planning, Monitoring, and Reflecting scales. Because such cross-loadings could not be theoretically justified, we discarded these six items. From the other scales, 11 items were trimmed to reduce the number of items while maintaining good coverage of the theoretical construct. The final CFA tested a model with 17 items which showed reasonable fit according to three of the four fit indices: χ^2 (113, $N = 162$) = 180.48, $p < .001$, RMSEA = .061 with a 90 % CI [.044, .077]., and CFI = .98, SRMR = .06. Composite reliability was good for all factors (range $\rho_c = .81-.91$). Based on the results from this pilot study, items from the scale Cognition Control were redeveloped. This resulted in the current instrument, the CEMOR questionnaire, consisting of 22 items in 5 scales (see Table 1 for the items).

5.2.4. Motivation and emotion scales

Children’s motivation was measured using translated items from the Children’s Perceived use of Self-Regulated Learning Inventory (CP-SRLI; Vandevelde et al., 2013). In their validation study, Vandevelde and

Table 1
Descriptive statistics of the CEMOR items and scales.

Scale	#	Item	*M*	*SD*	Range	Skewness	Kurtosis
Planning		Before I started the exercises, I asked myself:	2.69	1.04	1-5	.02	-.72
	Q1	"How can I make sure I don't make mistakes?"	2.46	1.16	1-5	.37	-.73
	Q2	"How can I make sure I feel good?"	2.73	1.26	1-5	.15	-1.04
	Q3	"How can I make sure I work hard?"	2.57	1.28	1-5	.28	-1.11
	Q4	"How can I make sure I try my best?"	3.01	1.32	1-5	-.09	-1.13
Monitoring		While making the exercises …	3.08	0.91	1-5	-.19	-.10
	Q5	I monitored whether I understood what I was doing	3.49	1.11	1-5	-.72	-.04
	Q6	I monitored my feelings	3.02	1.20	1-5	-.08	-.91
	Q7	I monitored how hard I was working	2.77	1.25	1-5	.11	-1.01
	Q8	I monitored whether I was trying hard enough	3.02	1.24	1-5	-.14	-.91
Cognition Control		Just now, I…	3.00	0.98	1-5	-.32	-.47
	Q9	Tried new ways to solve a difficult exercise	2.79	1.24	1-5	.03	-1.02
	Q10	Thought about a better way to solve a difficult task	2.87	1.25	1-5	-.09	-1.08
	Q11	Thought about other ways to solve a task	2.91	1.26	1-5	-.08	-1.06
	Q12	Tried to solve exercises in multiple ways	3.22	1.25	1-5	-.42	-.84
	Q13	Tried several approaches to solve a difficult task	3.23	1.21	1-5	-.51	-.75
Scale	#	Item	*M*	*SD*	Range	Skewness	Kurtosis
Emotion/Motivation Control		Just now, I…	2.67	0.89	1-5	.06	-.33
	Q14	Thought of things that make me happy	2.51	1.24	1-5	.42	-.79
	Q15	Tried to relax myself (breathing exercise, counting to ten, etc.)	2.48	1.25	1-5	.40	-.89
	Q16	Thought of ways to make the exercises more fun or challenging	2.44	1.20	1-5	.50	-.65
	Q17	Told myself: "Come on, you can do it!"	3.07	1.34	1-5	-.15	-1.14
	Q18	Told myself: "Start work and concentrate, these exercises are important."	2.86	1.27	1-5	.02	-.68
Reflecting		After doing the exercises, I asked myself:	2.59	1.00	1-5	.14	-1.01
	Q19	"How can I make even less mistakes next time?"	2.70	1.24	1-5	.25	-.86
	Q20	"How can I make sure I feel even better next time?"	2.59	1.17	1-5	.43	-.74
	Q21	"How can I make sure I work even harder next time?"	2.40	1.18	1-5	.17	-1.02
	Q22	"How can I make sure I do my best next time?"	2.66	1.25	1-5	.14	-1.01

Note. # shows how the items correspond to Fig. 1.

colleagues (2013) showed the CP-SRLI measures motivation with acceptable to good factorial validity (range factor loadings = .44−.91) and acceptable to good reliability (range Bentler's ρ = .74−.83). The CP-SRLI measures motivation in four scales (see Table 2 for example items): External, Introjected, Identified, and Intrinsic. Items were answered on a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree). Cronbach's alphas in the current study indicated the scales External (α = .61) and Introjected (α = .61) did not reliably measure their relative constructs. For the correlational analyses in this study, we therefore only used the data from the scales Identified (α = .83) and Intrinsic (α = .89), as these did show good reliability.

Perceived negative emotions were measured using translated items from the Achievement Emotions Questionnaire (AEQ; Pekrun et al., 2011). In their validation study, Pekrun and colleagues (2011) showed the AEQ measures emotions with reasonable factorial validity and good reliability (range Cronbach's alfa = .86.93). We used items from four scales: Anger, Anxiety, Boredom, and Hopelessness (see Table 2 for example items). Items were answered on a 5-point Likert scale, ranging from 1 (totally disagree) to 5 (totally agree). Cronbach's alphas for the current study indicated all scales measured their constructs with acceptable to good reliability (range Cronbach's alpha = .79.90).

5.3. Procedure

The math task and questionnaire were administered online via software program Qualtrics. After a short explanation video, the children had 25 min to work on the math task. Children were allowed to use scrap paper, but no calculator, and could skip a problem if they got stuck. After 25 min, children were automatically forwarded to the Emotion and Motivation scales and the CEMOR questionnaire. Children could complete these questionnaires at their own pace and could, if necessary, ask their teacher to rephrase items if they were uncertain about their meaning. Most children completed the questionnaires within 30 min. On average, the entire data collection took about 60 min to complete. As our online data collection forced children to answer all items, the response rate was 100 %. Participants did not receive a reward for participating in this study. Data were collected in the period of March - June 2021 in children's own classrooms. This research received ethical approval from the ethics committee of the first authors' university (project number: 2019-DP-11615).

Table 2
Example items and descriptive statistics of math difficulty and the perceived emotion and motivation scales.

Scale	Example item	No. items	*M*	*SD*	Range	Skewness	Kurtosis
Demand	I thought these exercises were hard.	5	3.39	0.71	1-5	.02	.18
Motivation	I have done my best on these exercises …						
External	… because others (my parents, the teacher, etc.) oblige me to do so.	3	1.93	0.72	1-5	.71	.25
Introjected	… because I would feel guilty if I didn't do my best.	4	2.42	0.83	1-5	.15	-.40
Identified	… because I want to learn new things.	4	3.63	0.87	1-5	-.60	.13
Intrinsic	… because I like doing math exercises	3	3.13	1.10	1-5	-.13	-.68
Emotions	While making these exercises …						
Anger	… I felt myself getting irritated.	3	2.05	0.97	1-5	.78	-.16
Anxiousness	… I was afraid I was making mistakes.	3	2.38	1.06	1-5	.47	-.62
Boredom	… I thought these exercises were fairly dull.	3	2.36	1.13	1-5	.70	-.25
Hopelessness	… I wanted to stop because I'm not going to understand it anyway.	3	1.99	0.98	1-5	1.02	.59

5.4. Data analysis

Children are most likely to engage in self-regulatory processes if they experience some difficulty (Cleary & Chen, 2009). To assess whether our math tasks were sufficiently challenging, we first analyzed children's perception of the math task's difficulty and their scores on the math problems. For perceived difficulty, we used the mean score on the Demand scale.

Internal factor structure of the CEMOR questionnaire was tested through confirmatory factor analysis (CFA), using *Mplus* (Version 7.31; Muthén & Muthén, 1998–2017). As our online data collection forced children to answer all items, there were no missing data. To account for the flatness of some of the SRL-items (kurtosis range −1.14 to −0.4; See Table 1), data were analyzed using robust maximum likelihood estimation (MLR). Goodness of fit was evaluated using Satorra-Bentler adjusted Chi-square (Satorra & Bentler, 2010), root mean square error of approximation (RMSEA), comparative fit index (CFI), and standardized root mean square residual (SRMR). The following values were used as indications of reasonable fit: SRMR and RMSEA close to or below .08, and CFI close to or greater than .95 (Hu & Bentler, 1999).

To ensure that the hypothesized internal factor structure fitted the data best, the a-priori hypothesized model was compared to several competing models. More specifically, the hypothesized 5-factor model was compared to a one-factor model (with all items loading on one general SRL factor), a hierarchical model (adding a second order variable of general SRL on top of the hypothesized 5 scales), a 4-factor model (merging the scales Cognition Control and Emotion/Motivation Control), and a 3-factor model grouping items based on what area they focus on (cognition, emotion, or motivation) instead of grouping them based on their respective phase (planning, monitoring, controlling, and reflecting). All alternative models were compared to the hypothesized 5-factor model using the Satorra-Bentler scaled Chi-Square test (Satorra & Bentler, 2010) and by comparing the fit-measures RMSEA, CFI, and SRMR.

To assess the quality of the CEMOR questionnaire, we looked at standardized factor loadings of the CEMOR items and correlations between the five CEMOR scales. Standardized factor loadings above.55 are seen as good indicators for their respective scale (Comrey & Lee, 1992). Moreover, correlations between the CEMOR scales should not be too low (>.32) as the scales measure related constructs, nor should they correlate too high (<.90) to justify being separate scales (Tabachnick & Fidell, 2014). Reliability of the CEMOR scales was evaluated by calculating composite reliability. Values above .70 are seen as acceptable and values above .80 are seen as indicators of good reliability (Brown, 1989).

To further investigate the validity of the CEMOR questionnaire, the mean scores on the SRL scales were correlated with children's performance on the math tasks and the mean scores on the Emotion and Motivation scales. Because intraclass correlation coefficients (ICC) indicated significant variance at the class level (range ICC on class level 0.03 – 0.11) for the Emotion and Motivation scales, we corrected for the multilevel structure in the data (children nested within classrooms) when calculating the correlations. Correlations below or around 0.10 are interpreted as small, around 0.25 as medium, and around or above 0.40 as large (Lipsey & Wilson, 2001).

6. Results

6.1. Descriptive statistics

Distributional properties of the CEMOR items and scales (see Table 1) were inspected to assess deviations from normality. Skewness levels did not exceed the threshold of ± 1.00 (range −0.72 to 0.50). In contrast, kurtosis of some items and scales did exceed the threshold of ± 1.00 (range −1.14 to −0.4) indicating these distributions were somewhat flat. The means of the CEMOR items and scales were mostly around the midpoint of the five-point Likert scale (M_{range} 2.40–3.49) with standard deviations and ranges indicating that the full range of answer options was represented. This indicates there were probably no bottom or ceiling effects.

Distributional properties of the Demand and the perceived Emotion and Motivation scales (see Table 2) were also inspected to assess deviations from normality and linearity. Except for the scale Hopelessness, no skewness levels passed the threshold of ± 1.00 (range −0.60 to 1.02). Similarly, no kurtosis levels passed the threshold of ± 1.00 (range −0.68 to 0.59). The means of the Demand and the perceived Emotion and Motivation scales were mostly around the midpoint of the five-point Likert scale (M_{range} 1.93–3.63) with standard deviations and ranges indicating a nice distribution over all answer options. This indicates there were no problems concerning bottom or ceiling effects.

6.2. Math task difficulty

On average, children answered 8.90 of the wizKID exercises correctly ($SD = 3.36$). Only two children answered more than 90 % of the wizKID math exercises correctly. Mean Demand was 3.34 ($SD = 0.70$). On average, children answered 6.93 of the wizSMART exercises correctly ($SD = 2.84$). No children answered more than 90 % of the exercises correctly. Children gave the wizSMART exercises a mean Demand score of 3.43 ($SD = 0.72$). The wizKID and wizSMART exercises can therefore be seen as adequately challenging.

6.3. Factor structure and composite reliability

After analyzing task difficulty, we ran a series of CFAs to evaluate whether the factors, when specified, corresponded to the hypothesized structure, and provided support for the internal validity and common factor structure of the CEMOR. Table 3 provides a comparison between the hypothesized 5-factor model and the alternative models. Results indicated that the 5-factor model fitted the data significantly better compared to the 1-factor, hierarchical, and 4-factor model, based on the Satorra-Bentler scaled chi-square difference tests. Moreover, the hypothesized model had better RMSEA, CFI, and SRMR scores compared to the alternative models.

Fig. 1 presents the standardized factor loadings of the CEMOR items and correlations between the CEMOR scales. Factor loadings (range .54 to.83) showed the items were good indicators of their respective scales. Moreover, the CEMOR scales were all significantly and positively correlated with each other (range .39−.64). These correlations were neither too high, nor too low, indicating that the scales indeed measured distinct yet related constructs. Furthermore, we found acceptable composite reliability for the scales Monitoring (ρ_c =.76) and Motivation/Emotion Control (ρ_c =.75), and good reliability for the scales Planning (ρ_c =.85), Cognition Control (ρ_c =.83), and Reflecting (ρ_c =.85).

6.4. Correlations

To further investigate the validity of the CEMOR, the mean scores on the SRL scales were correlated with children's performance on the math task, perceived emotions, and level of motivation (see Table 4). In contrast to our expectations, none of the SRL scales was significantly related to children's performance on the math task. Concerning Motivation, the scales Identified and Intrinsic were positively correlated to all CEMOR scales, showing small to medium correlations (range .14–.45). This indicates that having high identified or intrinsic motivation coincided with more self-regulated learning.

Regarding perceived emotions, Anger showed a significant but small positive correlation to the Planning and Reflecting scales (range .09−.11). Anxiousness was significantly correlated to all SRL scales, showing small to medium positive relations (range .13–.31). Boredom showed significant, small to medium negative correlations with all scales (range −.25 to −.10) except the CEMOR Planning scale. Last, Hopelessness was only significantly and negatively correlated with

Table 3
Satorra-Bentler scaled chi-square difference tests and fit measures for hypothesized and alternative models.

	Satorra-Bentler scaled χ^2 difference test			Fit measure				
	χ^2	Δdf	*p*	RMSEA [90 % CI]	CFI	SRMR	ΔCFI	ΔSRMR
Hypothesized model								
5-factor model	593.17			.060, [.055,.066]	.90	.05		
Alternative models								
1-factor model	1489.26	10	< .001	.106 [.101,.111]	.67	.09	-.23	.04
Hierarchical model	643.35	5	< .001	.063, [.057,.068]	.89	.06	-.11	.01
4-factor model	745.52	4	< .001	.070, [.065,.075]	.86	.07	-.14	.02
3-factor model*	1311.03			.099, [.094,.104]	.72	.09		

Note. Δdf, ΔCFI, and ΔSRMR were calculated by subtracting the respective value of the alternative model from the hypothesized 5-factor model. RMSEA = root mean square error of approximation; CI = confidence interval; CFI = comparative fit index; SRMR = standardized root mean square residual. 5-factor model = Planning, Monitoring, Cognition Control, Emotion/Motivation Control, Reflecting; 1-factor model = general SRL, Hierarchical model = Planning, Monitoring, Cognition Control, Emotion/Motivation Control, Reflecting as first order variables and general SRL factor as second order variable; 4-factor model = Planning, Monitoring, Cognition/ Emotion/Motivation Control, Reflecting; 3-factor model = Cognition Regulation, Emotion Regulation, Motivation Regulation.
* 3-factor model is not nested, so no Satorra-Bentler test conducted.

Cognition Control ($r = -.13$). In general, this indicates that children experiencing anger and anxiety tended to self-regulate more, whereas children experiencing boredom and hopelessness tended to self-regulate less during the math task.

7. Discussion

The aim of the current study was to develop and initially validate a new SRL instrument for upper elementary school. This had several reasons. First, most currently available questionnaires measure SRL as a general ability. However, children often find it hard to generalize about their SRL ability over multiple contexts (Endedijk et al., 2016), which makes these questionnaires less suitable for use in elementary education. Second, very little of these instruments cover the full concept of SRL, with motivational and emotional learning aspects particularly lacking (Koivuniemi et al., 2021). Third, SRL can be seen as a fluent ability that fluctuates between and sometimes even within situations (Boekaerts & Corno, 2005), which asks for a more fine-grained analysis of SRL in specific situations than is possible with currently available questionnaires. Moreover, alternatives like thinking aloud protocols and observations are often time-consuming and hard to administer in large samples. To alleviate these problems, we developed the CEMOR questionnaire, a task-specific instrument that measured the SRL that children applied during a task they just completed. In developing the CEMOR, we used a multistep procedure, including cognitive interviews, expert review panels, and pilot testing. The CEMOR questionnaire aligns with the model of Pintrich (2000), distinguishing between different phases (i.e., planning, monitoring, controlling, and reflecting) and areas of SRL (in this study we distinguished between cognition and emotion/motivation regulation).

The CEMOR was applied in a math context in this study. The results support the psychometric qualities of the CEMOR questionnaire in several ways. First, the means, standard deviations, and ranges of the CEMOR items show the answers spread around the center of the Likert scale. This indicates there were probably no bottom or ceiling effects. Moreover, the lack of ceiling effects could indicate that most of the participating children did not overestimate their SRL ability. This aligns with Veenman (2011a), who states that problems with memory reconstruction can be mitigated to some degree by reducing the delay between task performance and retrospective questioning to a minimum, and by making the retrospective questionnaire more task-specific. However, to make a substantiated claim about the degree of overestimations in the CEMOR results, a more thorough investigation is necessary. Future research could, for instance, administer both the CEMOR and another SRL questionnaire like the CP-SRLI to the same sample. By comparing these results to teacher reports of SRL or thinking aloud protocols, more information about the differences in accuracy between task-specific and more general questionnaires could be gathered.

Second, standardized factor loadings demonstrate that the items are good indicators of their respective scales. Of all CEMOR items, only 1 out of 22 items had a factor loading of .54 and therefore failed to meet the threshold of .55 for good indicators (Comrey & Lee, 1992). This exceeded our expectations based on results from factor analyses of similar SRL instruments. For comparison, the CP-SRLI counts 23 out of 75 items with a factor loading below the threshold of .55 (as reported in: Vandevelde et al., 2013). The higher factor loadings of the CEMOR could be due to the task-specific nature of our items. Some children find it hard to generalize about SRL ability across multiple times and situations (Endedijk et al., 2016), which can negatively impact how well items are able to measure their respective constructs. As our task-specific items do not ask for this generalization, this is a less common problem in our data.

Third, CFAs show that the CEMOR items can indeed best be grouped in five scales (i.e., Planning, Monitoring, Cognition Control, Emotion/ Motivation Control, and Reflecting), as our 5-factor model showed good model fit and fitted the data better compared to several alternative models. Especially relevant is that our 5-factor model outperformed the alternative 4-factor model. This supports the notion that cognition regulation and emotion/motivation regulation are indeed separate constructs that warrant measurement in separate scales for the control phase. Moreover, correlations between the five CEMOR scales were all significant, yet neither too high (>.9), nor too low (<.32), indicating that the scales indeed measure distinct yet related constructs. Moreover, composite reliability of the CEMOR scales was acceptable to good. These findings support the distinction between phases and areas of regulation based on the model by Pintrich (2000), showing that this distinction is indeed relevant and discernable when measuring task-specific SRL in elementary education.

In addition, we investigated the relationship between measured SRL ability and constructs that were hypothesized to be related to SRL: Children's performance on the math task, emotions, and motivation. In contrast to our hypothesis, we did not find a significant relationship between SRL and children's math performance. The absence of a relationship between SRL and math performance contrasts with earlier studies showing that the use of self-regulated learning strategies is a good predictor of the academic achievement of seventh graders (Pintrich & De Groot, 1990) and sixth graders (Ee et al., 2003). However, it aligns with several studies in secondary and higher education that found no significant relationship (Malpass et al., 1999; Sitzmann & Ely, 2011; Yoon, 2009). The fact that we did not find a significant relation with math performance could be due to the present study investigating a younger sample, and/or using a task-specific (event) measure compared to the general SRL (trait) measures used in other studies. Earlier research indicates that trait and event measures indeed explain different and unique amounts of variance in math achievement (Callan & Cleary, 2018). Follow-up research is warranted to better understand if and when

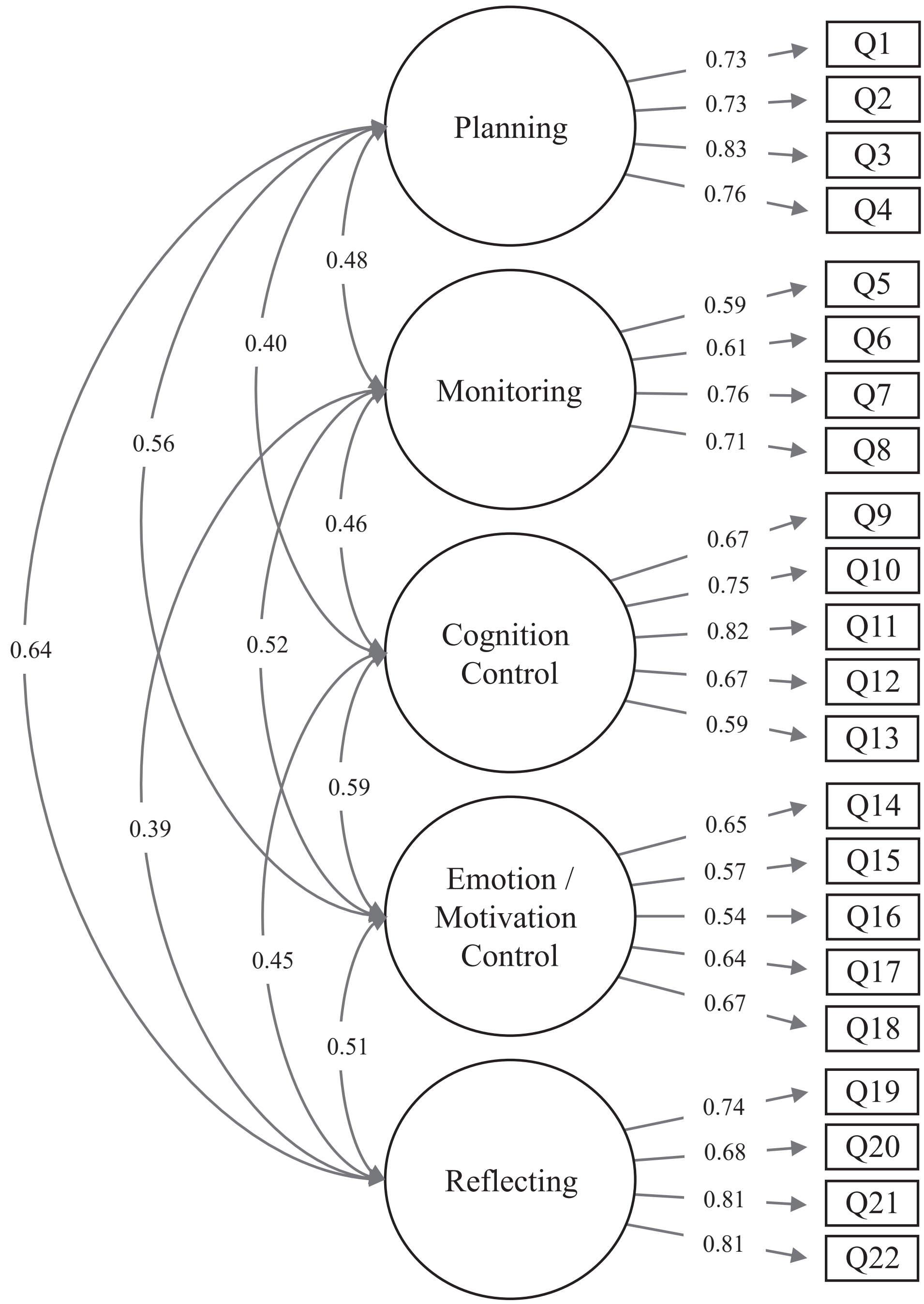

Fig. 1. Standardized Factor Loadings of CEMOR Items and Correlations between CEMOR Factors. *Note.* See Table 1 for the items corresponding to Q1-Q22. All reported correlations are significant ($p = <.001$).

Table 4
Correlations between CEMOR scales and Math Performance, Motivation, and Emotion scales.

Scale	Math Performance		Motivation		Emotion			
	wizKID[1]	wizSMART[2]	Identified	Intrinsic	Anger	Anxiousness	Boredom	Hopelessness
Planning	-.12	-.08	.38***	.20***	.11*	.31***	-.07	.07
Monitoring	-.03	.02	.45***	.25***	.01	.22***	-.15**	-.05
Cognition Control	.09	.06	.41***	.37***	-.06	.13**	-.25***	-.13**
Emotion/ Motivation Control	.01	-.05	.39***	.30***	-.01	.20***	-.18***	-.08
Reflecting	-.12	-.05	.32***	.14**	.09*	.30***	-.10*	.09

Note. [1] grade 3 and 4; [2] grade 5 and 6; * $p < .05$. ** $p < .01$. *** $p < 0.01$.

a relationship between SRL and math performance can be expected.

The relationship between SRL and emotions corresponded more closely to our expectations. Children's anger was positively related to their ability to plan and reflect on the task at hand, and their level of anxiousness was positively related to all SRL dimensions. These findings are consistent with earlier research defining these emotions as activating, meaning that a moderate amount of these emotions can motivate children to invest effort in order to overcome obstacles and avoid failure (Pekrun et al., 2002, 2007). In contrast to these positive relations, children's boredom was negatively related to all SRL dimensions except their level of planning. Moreover, children's feelings of hopelessness were negatively related to their level of cognition control. This aligns with earlier research describing these emotions as deactivating and interfering with task focus (Tyson et al., 2009).

Last, as expected, we found a positive relationship between both identified and intrinsic motivation and SRL. These results align with research showing that students with high identified and intrinsic motivation display deeper cognitive processing, more persistence, and better performance (Vansteenkiste et al., 2004). The fact that the relationships between SRL and emotions and motivation were observed in the expected directions, lends more credence to the CEMOR questionnaire.

7.1. Limitations

There are several limitations to this study that warrant discussion. First, due to the Covid pandemic, our researchers could not be present in the classrooms during data collection. To limit the possible problems this could evoke we decided to collect data in an online environment, including a short instruction film with all necessary information and a build-in timer that guaranteed children were automatically directed to the questionnaire after working on the math task for 25 min. This way we tried to limit the variation of data collection procedures between schools and classes. Nevertheless, because researchers were not present during data collection, we cannot be entirely sure data collection happened without irregularities.

Second, the generalizability of our sample warrants discussion. As this study was conducted with active consent, a subset of parents decided not to give their children consent for participation in this study, leading to a non-random selection of participants. Our sample, therefore, may not be representative of the ethnic diversity found among children in Dutch elementary education. For instance, in Dutch elementary education, 72 % of the children has an ethnic Dutch background (Centraal Bureau voor de Statistiek, 2022), whereas in our sample, 92 % of the children had a Dutch background. More research is warranted to investigate the measurement invariance of the CEMOR questionnaire over time and concerning specific student characteristics such as ethnic background, gender, and age.

7.2. Scientific and practical relevance

This study holds scientific relevance in multiple ways. First, it addresses the limitations of existing SRL questionnaires by developing and validating a task-specific instrument, the CEMOR questionnaire, specifically designed for upper elementary school children. This instrument allows for the fine-grained analysis of SRL in specific situations, which is often challenging with general SRL measures. By doing so, the study expands the measurement options for researchers interested in assessing SRL in elementary education. Secondly, this study aligns with the model of SRL by Pintrich (2000), distinguishing between different phases and areas of SRL, and provides empirical evidence supporting this distinction. This adds to the theoretical understanding of SRL and contributes to the ongoing refinement of conceptual frameworks in the field. Finally, the study explores the relationships between SRL, emotions, motivation, and academic performance, offering insights into the complex interplay among these variables.

The practical relevance of this study lies in its potential implications for educational practice and intervention. The development of the CEMOR questionnaire as a task-specific instrument provides teachers with a practical tool to assess and monitor students' SRL abilities, particularly in the context of specific tasks such as math. By utilizing this instrument, teachers can gain valuable insights into students' SRL processes and tailor instructional strategies accordingly. Furthermore, the positive relationship between identified and intrinsic motivation and SRL highlights the importance of fostering a supportive and motivating classroom climate to encourage children's self-regulation. Ultimately, the findings of this study have practical implications for the design of SRL education and interventions, allowing educators to develop evidence-based instructional practices and curricula that foster self-regulation skills in elementary education.

8. Future directions and conclusion

The useability and validity of the CEMOR questionnaire could be further explored by using the instrument with different tasks. The CEMOR items were phrased in a way to make this kind of follow-up research possible. The items do not refer specifically to math as the items only ask about children's self-regulation while making 'the exercises'. This allows future researchers to combine this same set of questions with a different task. For instance, it would be interesting to investigate to what extent this instrument is applicable to reading or scientific tasks. This approach would also allow researchers to investigate children's differences in SRL ability between different tasks. Moreover, future research could investigate whether the CEMOR questionnaire can be used with, for instance, even younger children or children from outside the Netherlands. Testing this new instrument with different tasks and more diverse samples will give extra information on the validity, measurement invariance, and useability of this instrument. Moreover, the validity could be further tested by relating the CEMOR results to other task-specific measures like thinking aloud protocols.

In conclusion, our study provides support for the CEMOR questionnaire as a valid and reliable instrument for measuring SRL in elementary education. The unique task-specific nature of the CEMOR questionnaire aligns with the view of SRL as situation specific, enabling a more fine-grained analysis of SRL within specific situations. Notably, the questionnaire avoids the need for children to generalize about their SRL ability across multiple instances, which may have contributed to the better understanding and accurate responses from children as young as 8 years old. This highlights the potential of this instrument for use with younger samples. Moreover, this instrument was designed to measure both the phases of SRL as well as both cognitive and emotion/motivation regulation. Therefore, the CEMOR questionnaire holds great promise for both scientific research and practical applications in evaluating children's SRL abilities across a broad spectrum.

Acknowledgements

We thank the schools, teachers, and students that participated in this project. We thank Merel Tuinenburg, Francien Bontekoe, Manon van Veen and Nora Winter for their help with data collection. We thank Maaike Zeguers for her help in the early stages of questionnaire development. We thank Suzanne Jak for troubleshooting and thinking along on the statistical analyses.

Author note

We have no conflicts of interest to disclose. This research was supported by the Research Priority Area YIELD, University of Amsterdam.

References

Alexander, P. A., Dinsmore, D. L., Parkinson, M. M., & Winters, F. I. (2015). Self-regulated learning in academic domains. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 393–407).

Ashcraft, M. H. (2002). Math anxiety: Personal, educational, and cognitive consequences. *Current Directions in Psychological Science, 11*(5), 181–185. https://doi.org/10.1111/1467-8721.00196

Bainbridge, L. (1979). Verbal reports as evidence of the process operator's knowledge. *International Journal of Man-Machine Studies, 11*(4), 411–436. https://doi.org/10.1016/S0020-7373(79)80035-8

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In Zimmerman, B.J., Schunk, D.H., (Eds.), Handbook of self-regulation of learning and performance (pp. 408–425). Routledge. https://doi.org/10.4324/9780203839010.ch26.

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology, 54*(2), 199–231. https://doi.org/10.1111/j.1464-0597.2005.00205.x

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction, 13*(3), 327–347. https://doi.org/10.1016/S0959-4752(02)00008-7

Brown, R. L. (1989). Using covariance modeling for estimating reliability on scales with ordered polytomous variables. *Educational and Psychological Measurement, 49*, 385–398.

Callan, G. L., & Cleary, T. J. (2018). Multidimensional assessment of self-regulated learning with middle school math students. *School Psychology Quarterly, 33*(1), 103–111. https://doi.org/10.1037/spq0000198

Centraal Bureau voor de Statistiek. (2022). *Leerlingen in (speciaal) basisonderwijs; migratieachtergrond, woonregio*. https://www.cbs.nl/nl-nl/cijfers/detail/83295NED?q=etniciteit basisonderwijs.

Cleary, T. J., & Chen, P. P. (2009). Self-regulation, motivation, and math achievement in middle school: Variations across grade level and math context. *Journal of School Psychology, 47*(5), 291–314. https://doi.org/10.1016/j.jsp.2009.04.002

Comrey, A.L., & Lee, H.B. (1992). A first course in factor analysis. Erlbaum.

Cornford, I. R. (2002). Learning-to-learn strategies as a basis for effective lifelong learning. *International Journal of Lifelong Education, 21*(4), 357–368. https://doi.org/10.1080/02601370210141020

Cromley, J., & Azevedo, R. (2011). Measuring strategy use in context with multiple-choice items. *Metacognition and Learning, 6*(2), 155–177. https://doi.org/10.1007/s11409-011-9070-z

De Corte, E., Mason, L., Depaepe, F., & Verschaffel, L. (2011). Self-regulation of mathematical knowledge and skills (Eds). In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance*. Routledge. https://doi.org/10.4324/9780203839010.ch10.

Dent, A. L., & Koenka, A. C. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review, 28*(3), 425–474. https://doi.org/10.1007/s10648-015-9320-8

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

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*(3), 231–264. https://doi.org/10.1007/s11409-008-9029-x

Dinsmore, D. L., Alexander, P. A., & Loughlin, S. M. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review, 20*(4), 391–409. https://doi.org/10.1007/s10648-008-9083-6

Ee, J., Moore, P. J., & Atputhasamy, L. (2003). High-achieving students: Their motivational goals, self-regulation and achievement and relationship to their teachers' goals and strategy-based instruction. *High Ability Studies, 24*, 23–39.

Endedijk, M. D., Brekelmans, M., Sleegers, P., & Vermunt, J. D. (2016). Measuring students' self-regulated learning in professional education: Bridging the gap between event and aptitude measurements. *Quality and Quantity, 50*(5), 2141–2164. https://doi.org/10.1007/s11135-015-0255-4

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin, 137*(2), 316–344. https://doi.org/10.1037/a0021663

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

Jitendra, A. K., Sczesniak, E., Griffin, C. C., & Deatline-Buchman, A. (2007). Mathematical word problem solving in third-grade classrooms. *Journal of Educational Research, 100*(5), 283–302. https://doi.org/10.3200/JOER.100.5.283-302

Jonassen, D. H. (2003). Designing research-based instruction for story problems. *Educational Psychology Review, 15*(3), 267–296. https://doi.org/10.1023/A:1024648217919

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., De Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M., & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist, 42*(3), 139–151. https://doi.org/10.1080/00461520701416231

Koivuniemi, M., Järvenoja, H., Järvelä, S., & Thomas, V. (2021). An overview of instruments for assessing and supporting elementary school students' self-regulated learning. *Learning: Research and Practice, 7*(2), 109–146. https://doi.org/10.1080/23735082.2020.1859123

Lipsey, M.W., & Wilson, D.B. (2001). Practical meta analysis. Thousand Oaks.

Malpass, J. R., O'Neil, H. F., & Hocevar, D. (1999). Self-regulation, goal orientation, self-efficacy, worry, and high-stakes math achievement for mathematically gifted high school students. *Roeper Review, 21*(4), 281–288. https://doi.org/10.1080/02783199909553976

McCann, E. J., & Turner, J. E. (2004). Increasing student learning through volitional control. *Teachers College Record, 106*(9), 1695–1714. https://doi.org/10.1111/j.1467-9620.2004.00401.x

Meneghetti, C., De Beni, R., & Cornoldi, C. (2007). Strategic knowledge and consistency in students with good and poor study skills. *European Journal of Cognitive Psychology, 19*(4–5), 628–649. https://doi.org/10.1080/09541440701325990

Midgley, C., Maehr, M.L., Hruda, L.Z., Anderman, E., Anderman, L., Freeman, K.E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M.J., Nelson, J., & Roeser, R. (2000). Manual for the Patterns of adaptive learning sciences (PALS). University of Michigan.

Moos, D. C., & Azevedo, R. (2008). Monitoring, planning, and self-efficacy during learning with hypermedia: The impact of conceptual scaffolds. *Computers in Human Behavior, 24*(4), 1686–1706. https://doi.org/10.1016/j.chb.2007.07.001

Muthén, L.K., & Muthén, B.O. (2017). Mplus user's guide. Eight Edition. Muthén & Muthén.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics, 102*(1), 9–28. https://doi.org/10.1007/s10649-019-09903-9

Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology, 8*, 1–28. https://doi.org/10.3389/fpsyg.2017.00422

Pekrun, R., Frenzel, A. C., Goetz, T., Perry, R. P., Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In P. A. Schutz, & R. Pekrun (Eds.), *Emotion in education* (pp. 13–36). Elsevier. https://doi.org/10.1016/B978-012372545-5/50003-4.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire. *Contemporary Educational Psychology, 36*, 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students̕self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105.

Perry, N. E., VandeKamp, K. O., Mercer, L. K., & Nordby, C. J. (2002). Investigating teacher-student interactions that foster self-regulated learning. *Educational Psychologist, 37*(1), 5–15. https://doi.org/10.1207/S15326985EP3701_2

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning (Eds). In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/B978-012109890-2/50043-3 (Eds).

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40. https://doi.org/10.1037//0022-0663.82.1.33

Poitras, E. G., & Lajoie, S. P. (2013). A domain-specific account of self-regulated learning: The cognitive and metacognitive activities involved in learning through historical inquiry. *Metacognition and Learning, 8*(3), 213–234. https://doi.org/10.1007/s11409-013-9104-9

Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research, 45*(3), 269–286. https://doi.org/10.1080/00313830120074206

Satorra, A., & Bentler, P. M. (2010). Ensuring possitiveness of the scaled chi-square test statistic. *Psychometrika, 75*(2), 243–248.

Schellings, G. (2011). Applying learning strategy questionnaires: Problems and possibilities. *Metacognition and Learning, 6*(2), 91–109. https://doi.org/10.1007/s11409-011-9069-5

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schuck, J. R., & Leahy, W. R. (1966). A comparison of verbal and non-verbal reports of fragmenting visual images. *Perception & Psychophysics, 1*(6), 191–192. https://doi.org/10.3758/BF03215783

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421–442. https://doi.org/10.1037/a0022777

Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*(1), 51–79. https://doi.org/10.1006/ceps.2001.1091

Tabachnick, B.G., & Fidell, L.S. (2014). Using multivariate statistics. Pearson.

Tyson, D. F., Linnenbrink-Garcia, L., & Hill, N. E. (2009). Regulating debilitating emotions in the context of performance: Achievement goal orientations, achievement-elicited emotions, and socialization contexts. *Human Development, 52*(6), 329–356. https://doi.org/10.1159/000242348

Usher, E. L., & Pajares, F. (2008). Self-efficacy for self-regulated learning: A validation study. *Educational and Psychological Measurement, 68*(3), 443–463. https://doi.org/10.1177/0013164407308475

Van Hout-Wolters, B., & Schellings, G. (2009). Measuring learning strategies. Kinds of measurement methods and their usefulness in educational research and practice. *Pedagogische Studien, 86*(1), 110–129.

Vandevelde, S., Van Keer, H., & Rosseel, Y. (2013). Measuring the complexity of upper primary school children's self-regulated learning: A multi-component approach. *Contemporary Educational Psychology, 38*(4), 407–425. https://doi.org/10.1016/j.cedpsych.2013.09.002

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Veenman, M. V. J. (2011a). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning, 6*(2), 205–211. https://doi.org/10.1007/s11409-011-9080-x

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

Wigfield, A., & Meece, J. L. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology, 80*(2), 210–216. https://doi.org/10.1037/0022-0663.80.2.210

Winne, P. H. (2017). The trajectory of scholarship about self-regulated learning. *Teachers College Record, 119*, 1–16.

Wolters, C. A. (2003). Regulation of motivation: Evaluating an underemphasized aspect of self-regulated learning. *Educational Psychologist, 38*(4), 189–205. https://doi.org/10.1207/S15326985EP3804_1

Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2006). The development and evaluation of procedures to assess child self-report item validity educational and psychological measurement. *Educational and Psychological Measurement, 66*(4), 687–700. https://doi.org/10.1177/0013164405282467

Yoon, C. H. (2009). Self-Regulated learning and instructional factors in the scientific inquiry of scientifically gifted korean middle school students. *Gifted Child Quarterly, 53*(3), 203–216. https://doi.org/10.1177/0016986209334961

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7.