Measurement of Executive Functioning Using a Playful Robot in Kindergarten

Corrie C. Urlings, Karien M. Coppens, and Lex Borghans

Macro, International and Labour Economics (MILE), School of Business and Economics, Maastricht University, Maastricht, MD, The Netherlands

ABSTRACT
We explored the potential of a robotics application in education as a measurement tool of child executive functioning skills. Sixty-five kindergarteners received assignments to go through a maze with a programmable robot, the Bee-Bot. Via observation we quantified how they solved these tasks. Their performance was successfully aggregated into a latent variable, which was used to predict the outcomes on standardized tasks that measure executive functioning. The latent variable significantly predicted performance on several tasks that measure problem-solving abilities, memory, visuospatial abilities and attention. It did not significantly predict pencil-and-paper tasks that measured visuospatial ability and nonverbal or design fluency. This study showed that it is possible to use a playful robotics task to obtain information on children’s ability. We recommend more research on using diverse robots in larger samples with different age groups to further explore the possibilities of robots as a test instrument.

KEYWORDS
Programming; robots; executive functioning; kindergarten

Introduction
In recent years robots have become available in kindergarten and primary school as a means to stimulate child development and enhance specific skills which are closely related to robots, such as programming and construction (Mitnik, Nussbaum, & Soto, 2008). These skills are related to specific cognitive abilities of children, executive functioning (EF), that allow higher order behavioral regulation on a daily basis. It consists of three core aspects, namely working memory, inhibitory control and cognitive shifting (Diamond, 2013; Miyake et al., 2000). Since measurements of EF are not easily performed in young children, that is, kindergarteners, the question is raised whether robots can be used to overcome some of the difficulties of traditional EF testing in kindergarteners.
History and the current use of robots in education

From computers and smartboards to specific games and programs designed for learning and cognitive stimulation, technology is present in the classroom even from a very early age onward. The newest trend in education is robotics, such as robot programming or construction. Since robots allow focus on technology and engineering, they are very suitable for STEM (Science, Technology, Engineering and Mathematics) education purposes. For example, according to Sullivan and Bers (2016), robotics provide opportunities for children to engage with technology and engineering concepts from an early age onward that beforehand were neglected or difficult to teach. In essence, technological applications and robots are employed in education as a means of teaching, stimulation, learning and possibly measurements.

Understanding executive functioning in kindergarteners

Understanding the cognitive ability of kindergarteners is of great importance, as it allows for comprehension of abilities and potential for tracking kindergarteners, to determine cognitive growth, potentially comparing children to their peers. EF is an umbrella term for cognitive processes that allow goal-directed behavior and self-regulation.

A multitude of tests has been developed to understand EF since studying EF is important. EF in childhood has been shown to have relations to a variety of outcomes later in life, such as theory of mind, self-concept, anti-social behavior and academic achievement (Hughes, 2011). However, the main components of EF were mostly based on research on adults. There is increasing consensus on the role of the individual EF components and the general model for younger children as young as three to four years of age (Hendry, Jones, & Charman, 2016). However, since young children often do not possess sufficient vocabulary or attention span, tasks most readily available for older children, testing children on EF (components) requires some creativity. For example, Carlson and Moses (2001) instructed children to point to white when “grass” was provided and point to green when “snow” was provided. Studies like this illustrate the playful or game-like aspects that are often incorporated into the testing moment. This makes sense, as it is generally accepted that children benefit from play and research has shown that children nurture their cognitive and overall development by playing (Goldstein, 2012; Vygotsky, 1967). Play is a crucial component in cognitive functioning (Ferrari, Robins, & Dautenhahn, 2009) and might therefore have a crucial role in measuring EF. Furthermore, Chaytor and Schmitter-Edgecombe (2003) found that the ecological validity of neuropsychological tests—used to measure executive functions, memory,
and attention processes—is mediocre at best when predicting everyday cognitive functioning skills. The authors stated that although the extensive test batteries may yield useful insights in clinical populations, the benefits might be lower when trying to understand healthy populations.

Now that more technological means have become available, this study will highlight the potential of using robots in relation to measurement issues in kindergarteners.

**Using robots to stimulate EF**

To fully understand the relation between kindergarteners and robotics, Sullivan, Kazakoff and Bers (2013) demonstrated that there were specific difficulties with implementing a robotics curriculum in pre-kindergarten classrooms, such as the need for one-on-one help from adults. Children were able to design, build and program a robot to perform a specific task after a one-week robotics intervention in the classroom. In an evaluation of a robotics curriculum, Bers, Flannery, Kazakoff and Sullivan (2014) employed the TangibleK Robotics Program paired with computer programming and robotic tools to engage kindergarteners in computational thinking, robotics, programming and problem solving. They concluded it was a viable technique and because the robot existed off-screen, there was more collaboration, shared material and practice of fine-motor skills. Sullivan and Bers (2016) showed that pre-kindergarten children were able to master the concepts used in programming a robot and that a 7 year old could add conditional statements to this as well. The authors highlighted the importance of working memory capacity for programming longer sequences and noted that younger children have a slower learning pace. De Michele, Demo and Siega (2008) investigated and described which developmental aspects were involved in a robotics project. Specifically for kindergarten and the lower primary grades, the Bee-Bot was used, followed by Scribbler by Parallax. The authors described that children developed several skills along the way, such as counting and logical thinking, solving topological problems, problem-solving and being acquainted with inquiry-based learning. Benitti (2012) performed a literature review, investigating whether robots were applied in a broader sense in education. The author found that 80% of the studies reviewed focused on physics and mathematics.

**Using robots to measure EF**

The game-like learning because of technology in the classroom, combined with the difficulties of testing kindergarteners, yields an interesting problem. A potential solution in playful assessment may lie in the application of
robotics in measuring ability. Like the creative tasks designed by researchers to understand EF, robots could collect data on the behavior of kindergarteners while they perform a task or play a game. We are particularly interested in the ability of robots to measure EF, as it is an established framework to understand cognitive abilities. Robotics tasks may be able to pick up variation between kindergarteners while they are conducting a fun task that lies close to their interests.

By envisioning a robotics task or game as not only a means to learn, but also as a data collection tool, we believe that technology provides opportunities to expand the understanding of ability. It also appears that in the younger age groups, the gain would be highest since creating measurements for this age group is quite intensive and is often a translation from comparable tests for older children, adolescents, or even adults. Furthermore, programming a robot that can move on the floor, rather than a computer-based task, might elicit more ecologically valid behavior from children, potentially yielding measures with higher ecological validity.

The Bee-Bot

The Bee-Bot has been found to be a very suitable robot for kindergarteners and can be programmed to navigate on a play mat (TTS Group Ltd). Sullivan et al. (2013), Highfield, Mulligan and Hedberg (2008) and Janka (2008) studied the interaction of children with a Bee-Bot in small sample sizes and made recommendations for implementing the technology in education and educational research. Sullivan et al. (2013) discussed the Bee-Bot in particular as a suitable robot for pre-kindergarten children. The study of Highfield et al. (2008) described motivated behavior in the children as they were interacting with the robot. The children employed a variety of problem-solving strategies, such as trial and error, recall of prior knowledge and investigating multiple solutions. Janka (2008) described that the Bee-Bot mostly elicited enthusiastic behavior in children, reinforcing the attractiveness for open-ended activities. The authors stressed the role of the teacher, as guidance is needed and groups of children working on a task should be small.

Recently, a study investigated the role of programming the Bee-Bot as an intervention to increase EF in children of 5–6 years old (Di Lieto et al., 2017). The children participated in three neuropsychological assessments during pre- and post-test. Outcome measures used in this study were scores of tests measuring EF (i.e., inhibition and working memory), visuo-spatial skills and attention. The authors concluded that EF, in particular working memory and inhibition, improved after the Bee-Bot intervention. Based on these studies, we concluded that the Bee-Bot is particularly useful
for working with kindergarteners, provided there is sufficient one-to-one or small-group interaction with a supervising adult, since its design will allow a more playful procedure to unfold.

**Current study**

The increased prevalence of robots in classrooms yields interesting opportunities for measurement issues in kindergarteners, since insight in EF of young children is important but difficult to operationalize. In order to answer the question if a playful robot task can yield information on EF, we allowed kindergarteners to perform a Bee-Bot task and systematically quantified their behavior and outcomes. We related these outcomes to the children’s outcomes on several tests that measure EF. In order to gain sufficient insight into the relation between outcomes on the Bee-Bot task and EF, we included a range of tasks that cover a wider spectrum of EF abilities. This approach has two advantages. The first is the limited number of validated EF tasks available for children in this age group. Second, including measures that have a slightly wider scope of skills can provide insights into the scope of the measurement. The different measures included in the EF test battery measure planning ability and memory, as well as visuospatial abilities, verbal and design fluency and a nonverbal intelligence measure. The scores on these tasks were compared to observables from the Bee-Bot task, such as the time they needed to think before acting (thinking time), the time needed to solve the task (execution time) and the number of errors they made. The EF tasks included were scored according to the principle that a higher score corresponds to higher ability. In some of the other tasks included in this study, such as a planning task or a visuospatial ability task, children have to provide the correct answer within a specific time frame in order to be scored as correct. We used the same rationale for solving the Bee-Bot task and thus hypothesized that the variables of the robotics task have a negative relation with outcomes on the EF tasks. Making exact predictions on the relationships between the Bee-Bot task outcomes and the EF test battery is difficult. We will therefore discuss the implications of the specific observations made during play with robotics, in relation to the EF tasks we included.

In this study, we examined the link between performance on Bee-Bot tasks and performance on standardized EF tasks. Our main research question was What EF are measured when children program a Bee-Bot? We first determined whether gender and age influenced the performance on both the Bee-Bot and the EF tasks. Next, we modeled the Bee-Bot data (thinking time, execution time and errors made) into a latent variable, which we then related to the EF tasks to answer the research question.
These EF tasks measure multiple EF such as planning, memory and working memory, attention, reasoning and problem solving, processing speed and flexibility. Since we aimed to gain a good understanding of the relation of EF, we included a broad spectrum of EF tests.

**Method**

**Participants**

We tested 65 children in kindergarten classes of elementary schools in the southern region of the Netherlands. Parents received an information letter and a consent form in Dutch. Participation was voluntary. Several exclusion criteria were formulated: Children were not to have doubled a grade, at least one of their parents must be fluent in Dutch in order to understand the consent and information forms and the child should not be diagnosed with a (developmental) disorder that could hinder his or her performance in some way (e.g., epilepsy, autism).

**Tasks, instruments and measurements**

*Bee-Bot*. The Bee-Bot looks like a bee with seven colorful buttons on top (Figure 1a). By pushing the buttons, a sequence can be programmed. Two buttons serve a forward and backward motion, two serve a left or right rotation. There is a ‘go’ button to execute the demand; a ‘clear’ button to reset the memory. The ‘pause’ button was not used in this study. The Bee-Bot communicates via sounds and lights to confirm the commands. A wooden maze with changeable walls was used as a play mat (Figure 1(b)).
**Introductory Bee-Bot lesson.** Children participated in the introductory class prior to testing, which was provided by the researcher in groups with a maximum five children for about 15 minutes to ensure an equal baseline of knowledge on the Bee-Bot. Each child programmed the Bee-Bot to go to the first letter of his or her own name on a letter mat.

**Bee-Bot task.** We developed three tasks for the children to perform in increasing difficulty for testing (Figure 1c). The assignment was to direct the Bee-Bot to the end of the maze, starting from the same starting point. After a reminder of the introductory class, the child was provided with the first assignment. Children were seated on a pillow on the floor in front of the maze, the researcher sitting on another pillow next to the child. There was no time limit and children could retry once per assignment if needed.

Using a stopwatch and scoring form, the researcher observed specific behaviors of the children while they were engaging with the Bee-Bot. The first variable registered was thinking time in seconds, for which measurement started after the instruction had been given and stopped when the child pressed the first button. The second variable was execution time, which is the time needed to program the Bee-Bot in seconds. This measurement started when the child pressed the first button and continued until the Bee-Bot reached the end position. The third variable was the number of errors children made during each assignment. The following errors were registered: pressing a button too often or not often enough and pressing the wrong button.

**Raven’s Coloured Progressive Matrices (RCPM).** On each page of a test booklet, a patterned rectangle shape was depicted with a piece missing from the pattern. Children had to indicate the right piece that fit the pattern from six options depicted on the same page. The booklet contained 36 items, divided into three sets (A, Ab and B). The Raven’s Coloured Progressive Matrices (RCPM) is often used to estimate general intelligence of children (Raven, Raven, & Court, 1998). Lezak, Howieson, Loring and Fischer (2004) stated that the task also measured visuospatial reasoning and problem solving. Raven et al. (1998) summarized that a low initial retest reliability of .65 for children under the age of 7 was found, but that subsequent studies yielded satisfactory reliability when assessed by split-half or retest methods. The score was the total number of correct items (maximum 36).

**Tower of London.** The Tower of London (ToL) task stems from the NEPSY test battery (Korkman, Kirk, & Kemp, 1998). Designed for children from
5–12 years old, the test consists of moving three balls on three pegs in such a way that it is a representation of a picture shown to a child. An answer was correct if the puzzle was solved with a specific number of steps within a specific time limit. Children were allowed to move only one ball at a time. At most 20 items are administered, unless four successive items were incorrect. The mean split-half reliability in a study performed by Korkman et al. (1998) was .82. The ToL task measures planning abilities and problem solving (Unterrainer et al., 2004). Studies with adults indicated also that the ToL loads on working memory, response inhibition and visuospatial memory (Carlin et al., 2000; Phillips, 1999; Welsh, Revilla, Strongin, & Kepler, 2000), but no such studies with children could be retrieved. Raw scores were transformed into scaled scores (maximum score 19).

**Five-point test.** This task measures the production of novel designs under time constraints and can be seen as a figural fluency test originally designed by Regard, Strauss and Knapp (1982). Children received a sheet of paper containing 40 boxes, containing five dots each. They had to draw as many unique designs as possible by connecting at least two dots with straight lines. Two sample drawings were made by the researcher, followed by at least two sample drawings made by the child before the test. Each drawing was scored as either correct, false (e.g., using curved lines), or double. In order to limit cognitive strain on the children, we used a one-minute restraint for this task. Data on reliability is sparse and mostly focuses on adults and normal populations (Strauss, Sherman, & Spreen, 2006). Tucha, Aschenbrenner, Koerts and Lange (2012) reported strong relations between performance on the five-point test (FPT), processing speed and mental flexibility, as well as medium relations with figural short-term memory, problem solving and inhibition. We used the total number of correctly drawn figures.

**Verbal fluency task.** This task is an adaptation of the NEPSY version (Korkman et al., 1998). The researcher explained that children had to name as many animals as possible within a one-minute timeframe. The researcher then gave two examples, namely, cat and dog, to ensure that children knew what was expected. Children’s responses were noted by the researcher and were later scored as correct animals, false animals (e.g., fantasy animals such as dragons) and doubles. Ardila, Ostrosky-Solis and Bernal (2006) showed verbal fluency task (VFT) correlates with immediate and delayed verbal memory. We used the total number of correctly named animals.
(Reverse) Digit span. This task is a subtest of the WISC-III (Kort et al., 2002). The researcher read numbers aloud and children had to repeat them aloud. Starting with two numbers in the first round, difficulty increased per two test items, by adding another number. After the “forward” items, children had to repeat number sequences backward. In both cases, testing stopped when the child could not repeat the number sequence successfully in both items of the same difficulty level. There was no time constraint for the child per item. Since part of our sample was slightly younger than the target age for the WISC-III, we treated the total correctly repeated sequences of the digit span (max. 16) and the reverse digit span (max. 14) as separate scores. A reverse digit span (RDS/DS) task has been successfully used with children aged 4 to 5 years (Alloway, 2007), but some report too little variability in a similar age range (Bull, Espy, & Wiebe, 2008). Lezak et al. (2004) reported that DS relies on attention, whereas RDS relies on mental tracking or working memory in adults.

Mazes. The mazes (MA) task is a subtest of the WPPSI-R (Vander Steene, & Bos, 1997). Children had to solve several mazes on paper with increasing difficulty. Mazes were scored as correct if solved within a certain time limit and with less errors made than a predetermined maximum amount (e.g., entering a blind alley or crossing a wall). Number of mistakes allowed increased with difficulty level of the mazes, but lead to a deduction in the score received for each particular maze. The test is based on the Porteus maze test (Porteus, 1959), which has been linked to planning ability (Carlozzi, 2011). The raw score was transformed into a scaled score (max. 19).

Procedure

Testing was divided over two sessions. During the first session the RCPM, FPT and MA were administered. During the second session the Bee-Bot task, ToL, VFT and DS were conducted. Age at the time of the second testing was included in the analysis, as well as gender of the child. Tests were administered in the school, in a separate room to accommodate one-to-one testing.

Analysis

For our analysis, we use the following variables. We included gender and age to understand whether these may confound the relationship between Bee-Bot and EF outcomes. Regarding Bee-Bot variables, we included for
each of the three assignments thinking time, execution time and number of errors. For EF tasks we included one variable for each task; namely, ToL, DS, RDS, VFT, RCPM, FPT and MA. We then used Mplus 7 to build a model in which we created a latent Bee-Bot variable, based on factor loadings of the included outcome variables of the Bee-Bot task. The latent variable predicted the performance on the various tests included in the test battery. All tests were included in one model.

**Results**

**Descriptive statistics**

Our sample consisted of 65 children, consisting of 28 boys and 37 girls (43/57%). The age of the children was computed in months. The mean age was 72.57 months (6 years), with a standard deviation of 4.5 months. T tests revealed there was no significant gender difference on the Bee-Bot outcomes, nor on the tasks in the EF test battery, except for the RCPM, t(63) = –3.40, p < .01, with girls outperforming boys. Simple regression analysis revealed that age in months did not significantly predict any of the tests from the EF test battery, nor any of the Bee-Bot outcomes.

Tables 1 and 2 show descriptive statistics and frequencies for the Bee-Bot task and the EF test battery. Execution time increased as the

---

**Table 1.** Descriptive information on thinking time and execution time of the three Bee-Bot tasks.

|               | N  | M     | SD  | Min. | Max. |
|---------------|----|-------|-----|------|------|
| Thinking time 1 | 65 | 4.91  | 6.75| 0    | 45.29|
| Thinking time 2 | 65 | 3.58  | 3.50| 0    | 17.39|
| Thinking time 3 | 65 | 4.95  | 4.54| 0    | 18.32|
| Execution time 1 | 65 | 18.41 | 7.31| 0    | 49.73|
| Execution time 2 | 65 | 53.42 | 18.02| 0    | 115.95|
| Execution time 3 | 65 | 12.89 | 1.79| 8    | 16   |
| ToL           | 65 | 6.62  | 1.77| 4    | 11   |
| DS            | 65 | 2.63  | 1.26| 0    | 5    |
| RDS           | 65 | 10.94 | 3.77| 3    | 21   |
| VFT<sup>a</sup> | 65 | 22.92 | 4.45| 9    | 34   |
| RCPM          | 65 | 6.79  | 2.36| 1    | 11   |
| MA            | 63 | 10.35 | 2.19| 7    | 18   |

Note. *Number of correct answers.

**Table 2.** Frequencies of errors during the three Bee-Bot tasks.

|               | Task 1 | Task 2 | Task 3 |
|---------------|--------|--------|--------|
| 0 errors      | 61     | 18     | 35     |
| 1 error       | 3      | 28     | 8      |
| 2 errors      | 1      | 18     | 15     |
| 3 errors      | 0      | 1      | 7      |
| Total         | 65     | 65     | 65     |

Note: The frequencies of the number of errors will be labeled as Errors 1, Errors 2 and Errors 3 in all following analysis, indicating the number of mistakes made by children per Bee-Bot assignment.
level of difficulty of the tasks increased, but thinking time remained stable over the different tasks. The number of errors also increased as the assignments became more difficult. As can be seen, range in the RDS was rather low. Therefore, only DS was used in the analysis further on. As can be seen in Table 3, thinking time variables correlated strongly. The highest correlation between two Bee-Bot measurements can be considered moderate (Taylor, 1990), at \( r(63) = .54, p < .01 \), between the number of errors during the second and third maze assignment. Table 4 shows the correlations between the Bee-Bot measures and the EF tasks. Execution time for the first assignment correlated significantly with the ToL and RCPM and slightly with the VFT. The number of errors on the second and third assignment correlated significantly with ToL, DS, VFT and the RCPM.

### Confirmatory factor analysis

In the confirmatory factor analysis, the correlations between the residuals of the regression predicting the outcomes on the test battery were set to 0, since we did not want to include the extent to which the EF tasks predicted each other. The outcomes on the Bee-Bot tasks were loaded onto a latent variable called 'Bee-Bot'. The first factor loading on the latent variable was

| Measure | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---------|---|---|---|---|---|---|---|---|---|
| T1      | – | – | – | – | – | – | – | – | – |
| T2      | .21* | – | – | – | – | – | – | – | – |
| T3      | .33*** | .49**** | – | – | – | – | – | – | – |
| Ex1     | .30** | .35*** | .22* | – | – | – | – | – | – |
| Ex2     | .16 | –.05 | –.10 | –.09 | – | – | – | – | – |
| Ex3     | .05 | .03 | –.07 | .00 | .11 | – | – | – | – |
| Er1     | .10 | –.05 | –.11 | .26** | –.07 | .01 | – | – | – |
| Er2     | .26*** | .10 | .21* | .45**** | –.10 | .08 | .18 | – | – |
| Er3     | .42*** | .08 | .09 | .50*** | .10 | .07 | .15 | .54*** | – |

Note. ***: \( p < .01 \), **: \( p < .05 \), *: \( p < .1 \).

The abbreviations T1, T2 and T3 refer to Thinking time 1, 2 and 3. Ex1, Ex2 and Ex3 refer to Execution time 1, 2 and 3. Er1, Er2 and Er3 refer to Errors 1, Errors 2 and Errors 3.

| Measure | ToL | DS | FPT | MA | VFT | RCPM |
|---------|-----|----|-----|----|-----|------|
| Thinking 1 | –.12 | –.14 | .09 | –.05 | –.02 | –.04 |
| Thinking 2 | –.06 | –.01 | –.09 | .03 | –.07 | –.24 |
| Thinking 3 | .02 | .01 | –.04 | –.03 | –.12 | –.17 |
| Execute 1 | –.33*** | –.19 | –.01 | –.01 | –.10 | –.28** |
| Execute 2 | –.16 | –.06 | –.07 | .02 | –.21* | .03 |
| Execute 3 | –.20 | –.08 | –.05 | –.02 | –.13 | .02 |
| Errors 1 | –.23* | –.17 | –.02 | –.18 | –.06 | –.05 |
| Errors 2 | –.43*** | –.45**** | –.04 | –.17 | –.29** | –.45*** |
| Errors 3 | –.45*** | –.36*** | –.09 | –.01 | –.41*** | –.21* |

Note. ***: \( p < .01 \), **: \( p < .05 \), *: \( p < .1 \).
freed and the latent variable was constrained to a mean of 0 and a variance of 1. We simultaneously regressed the latent 'Bee-Bot' variable on the outcomes of children in the EF test battery. The residuals of the regressions had a fixed correlation of 0. The number of participants was 65, with 2 missing values on the mazes variable. Mplus controls for missing values in modeling the data. The chi square of the initial factor model was not significant at a .05 level, with $\chi^2 (90, N=65) = 110.84, p > .05$. RMSEA was .06, the CFI/TLI values were .84/.81. We adjusted the model based on modification indices reported, so a correlation was added between the residuals of thinking time of the second and third assignment (M.I. = 14.06; E.P.C. = 7; StdYX E.P.C. = .47). The standardized factor loadings of the adjusted model can be found in Table 5 and the standardized coefficients in Table 6. The ToL, DS, VFT and RCPM had highly significant, negative coefficients. The FPT and MA task had no significant coefficient, but the coefficients were negative. The standardized correlation between thinking time 2 and 3 was highly significant, $r(63) = .47, p < .01$. Figure 2 shows a graphical depiction of the model. The fit of this model was good, $\chi^2(89, N=65) = 95.16, p > .1$, RMSEA = .03, CFI/TLI = .95/.94. We also ran the model individually, in which the latent 'Bee-Bot' variable predicted the tests from the EF test battery individually. The models were constructed as the all-inclusive model. From these separate models, again the ToL, DS, VFT and RCPM had highly significant, negative regression coefficients. The FPT and MA task had no significant coefficient, but the sign was also negative.

| Table 5. Standardized factor loadings of Bee-Bot variables on latent Bee-Bot variable. |
|------------------|-------|-------|-----------------|-----------------|
|                  | Estimate | SE   | Est./SE | p Value (two-tailed) |
| Thinking time 1  | .38    | .12  | 3.08    | .00              |
| Thinking time 2  | .20    | .13  | 1.48    | .14              |
| Thinking time 3  | .21    | .13  | 1.55    | .11              |
| Execution time 1 | .59    | .10  | 6.08    | .00              |
| Execution time 2 | .04    | .14  | .29     | .77              |
| Execution time 3 | .12    | .14  | .88     | .38              |
| Errors 1         | .26    | .13  | 2.01    | .05              |
| Errors 2         | .77    | .07  | 1.52    | .00              |
| Errors 3         | .73    | .08  | 9.17    | .00              |

| Table 6. Standardized coefficients of regression of the EF test battery outcomes on latent Bee-Bot variable. |
|------------------|-------|-------|-----------------|-----------------|
|                  | Estimate | SE   | Est./SE | p Value (two-tailed) |
| ToL              | -.58    | .10  | -6.00   | .00              |
| DS               | -.55    | .10  | -5.34   | .00              |
| VFT (correct)    | -.39    | .12  | -3.22   | .00              |
| RCPM             | -.49    | .11  | -4.35   | .00              |
| FPT (correct)    | -.12    | .14  | -.89    | .38              |
| MA               | -.16    | .14  | -1.15   | .25              |
Discussion

Robots are used in education mostly to enhance learning in children, specifically related to programming, technology and engineering skills. In this study, we explored the potential of robots to measure EF in kindergarteners. This is valuable since measuring ability in kindergarteners is difficult. We used a Bee-Bot, which is a programmable robot suitable for kindergarteners, as a measurement instrument for EF. Behaviors of the children were observed while they solved a task using the robot, in which the robot had to be navigated through several mazes. We quantified the observations to variables; namely, the time children thought about the assignment before they acted (i.e., thinking time), the time they needed to solve the assignment (i.e., execution time) and how many mistakes they made (i.e., number of errors). These variables were loaded into a latent Bee-Bot variable, which was used to predict the outcomes of the children on other tasks included in the test battery. We measured EF via the use of the Raven Coloured Progressive Matrices, Tower of London, Digit Span, Mazes, Verbal fluency...
task and Five-point test which were included in the test battery. Each task relies on one or more EF to be performed, such as planning (ToL, MA), memory (ToL, VFT, FPT), and working memory (ToL, RDS), attention (DS), reasoning and problem solving (RCPM, ToL, FPT), processing speed and flexibility (FPT) as well as inhibition (ToL, FPT).

**Interpretation of results**

The research question of the current study was which EF can be measured when children program a Bee-Bot. We gained insight into how different types of information stemming from observations contribute to a latent ‘Bee-Bot’ variable based on the factor loadings. In constructing the latent variable, the number of errors children made on more difficult tasks loaded onto the latent factor, as well as the amount of time (both thinking and execution time) needed for the first assignment. Given that all Bee-Bot variables contributed in a positive manner to the latent variable, it makes sense that the latent variable was negatively associated with outcomes on the EF tasks. That is, children with longer thinking and execution time and more errors, scored significantly lower on the ToL, DS, VFT and RCPM. The hypothesis that there is a negative relation between programming the robotic toy and particular aspects of EF was confirmed. The latent variable is a strong predictor for performance on the RCPM, ToL, VFT and DS. These tasks rely on problem solving (ToL, RCPM), memory (ToL, VFT), visuospatial reasoning and visuospatial memory (RCPM, ToL) and attention (DS). However, the latent variable was not a strong predictor for performance on the FPT and MA. These two tasks also rely on processes such as memory and problem solving, but also on processing speed, flexibility, inhibition (FPT) and planning (MA).

The different EF tests were included all together in the model. Miyake et al. (2000) argued that different aspects of EF are separable, yet moderately correlated constructs. Our test battery encompassed a broader spectrum of EF. In order to get the full picture of the relation between what the latent variable measures and the various aspects of EF and related tasks, the choice to include them all simultaneously in one model seems justified given the moderate correlations we found between several of the tasks in the test battery, thus providing a less clear picture.

**Implications of the findings**

The results show with which EF the latent ‘Bee Bot’ variable corresponds the most, thus providing insight into the relationship between programming a Bee-Bot and EF. The complexity of programming a task could be a mixture
of skills picked up by our test battery. Skills needed for programming (robots) have been labeled as computational thinking skills in some studies, which according to Wing (2006), can be defined as “taking an approach to solving problems, designing systems and understanding human behavior that draws on concepts fundamental to computing” (p. 3717). Román-González, Pérez-González and Jiménez-Fernández (2017) compared the Computational Thinking test (González, 2015) to other standardized psychological tests and concluded that the results correlated significantly with spatial ability, reasoning ability and strongest with problem-solving ability. We cannot state that the Bee-Bot task can represent the full spectrum of the computational skill set. However, aggregated performance on a Bee-Bot task in our study significantly predicted problem solving, memory, visuospatial abilities and attention. Furthermore, the RCPM in our study is related to reasoning ability, which may be considered overlap as well. Of course, this is not conclusive, since we did not measure the exact abilities in an identical way.

Regarding the regression coefficients for each task in the test battery, Di Lieto et al. (2017) showed a significant positive relationship between EF and a short Bee-Bot intervention training in the same age group as our current study, more specifically visuospatial working memory and inhibition skills. Our study showed a significant relation between the Bee-Bot task and a verbal memory task, which stresses the importance of memory for solving a robotics task. We also found a clear relation with visuospatial abilities and memory. Di Lieto et al. (2017) found that the Bee-Bot intervention did not increase visuospatial abilities in itself, but only visuospatial working memory. Therefore, the importance of visuospatial abilities in relation to (working) memory processes are stressed. Furthermore, Kazakoff, Sullivan and Bers (2013) found that a robotics and programming workshop for children over 4.5 years old significantly improved their sequencing (i.e., planning) ability. These findings are therefore somewhat consistent, since we also found that the latent Bee-Bot variable could explain a significant portion of a task measuring planning ability (ToL), but could not be confirmed with another task measuring the same EF (MA).

Nevertheless, the latent Bee-Bot variable was able to predict verbal fluency ability, but not figural fluency. The VFT does require memory ability, since answers have to be retrieved from long-term memory and children have to keep track of answers they already provided. However, in the FPT, the child does not need to have specific knowledge to be able to perform the task, but he or she does need memory ability to track answers for uniqueness. It may also be possible that the FPT was confounded by drawing ability. Potentially holding the pencil and drawing puts a strain on a child in this age group, leaving insufficient capacity behind for the task itself. The same might hold for MA.
The findings of our study are relevant for researchers working on innovative ways to measure EF or who work with programming abilities in young children. The ‘Bee-Bot’ is available for purchase and can easily be used for new studies. Furthermore, it is also easy to use by educators. The understanding of the EF skills related to kindergartener’s working with the Bee-Bot can help educators understand when and why a Bee-Bot can be used in the classroom. For instance, when doing assignments with the Bee-Bot on a mathematics or letter mat, the educator can be informed as well on which EF skills are related to the children’s performance.

 Limitations of the study

A bias may be introduced as to which children can join in, since participation was voluntary. The sample size was relatively small and some children may be more experienced with the nature of the task that was given to them than others. We tried to circumvent this problem by providing all children with an introductory class on the use of the robot. Taken together, it could be that the sample is not representative of the population. Lastly, we did not include an inhibition task in this study, which we recommend is included in future research as well.

 Conclusion

This study showed that letting kindergarteners solve mazes using a Bee-Bot can yield interesting insights into the memory ability, non-verbal ability, verbal fluency and planning ability of kindergarteners. With the exception of verbal fluency, these results are in agreement with studies that examine whether a robotics or programming intervention leads to an improvement of these specific skills. The difference is that our study does not focus on increasing skills, but on measuring them in a more playful way using robots, contributing to our understanding of whether robotics can be used to measure EF. The potential is there to not only use on-screen or computer-based assessments, but to attempt new perspectives that are possible due to evolving technologies. Sensors in robots could replace an observing researcher who is quantifying behavior visually, for example. Ideally, there are several robotic toys available, which require different types of programming and differing levels of difficulty or abstraction (e.g., visual images versus code) to measure different aspects and levels of EF. Together with the maze task performed using the Bee-Bot, a comprehensive view on EF could potentially be acquired using programmable robots. Hence, more research in this direction could increase the understanding of EF or general ability and the ease with which kindergarteners could be tested.
Acknowledgments

This study was funded by the Human Enhancement and Learning (HEaL) subsidy of Maastricht University, which aims to stimulate interdisciplinary and innovative research on human learning. The Ethical Review Committee of Psychology and Neuroscience at Maastricht University (ERCPN-175_02_01_2017) approved this study.

References

Alloway, T. P. (2007). Working memory, reading, and mathematical skills in children with developmental coordination disorder. *Journal of Experimental Child Psychology, 96*(1), 20–36.

Ardila, A., Ostrosky-Solis, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of semantic verbal fluency (ANIMALS). *International Journal of Psychology, 41*(5), 324–332. doi:10.1080/00207590500345542

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education, 58*(3), 978–988. doi:10.1016/j.compedu.2011.10.006

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. *Computers & Education, 72*, 145–157. doi:10.1016/j.compedu.2013.10.020

Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*(3), 205–228. doi:10.1080/87565640801982312

Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia, 38*(5), 655–665. doi:10.1016/S0028-3932(99)00102-5

Carlozzi, N. E. (2011). Porteus maze. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of clinical neuropsychology* (pp. 1964–1966). New York, NY: Springer.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children’s theory of mind. *Child Development, 72*(4), 1032–1053. doi:10.1111/1467-8624.00333

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review, 13*(4), 181–197. doi:10.1023/B:NERV.0000009483.91468.fb

De Michele, M. S., Demo, G. B., & Siega, S. (2008, November). A Piedmont schoolnet for a K-12 mini-robots programming project: Experiences in primary schools. Paper presented at the International Conference on Simulation, Modeling and Programming for Autonomous Robots, Venice, Italy.

Di Lieto, M. C., Inguglietto, E., Castro, E., Cecchi, F., Cioni, G., Dell’Omo, M., … Dario, P. (2017). Educational robotics intervention on executive functions in preschool children: A pilot study. *Computers in Human Behavior, 71*, 16–23. doi:10.1016/j.chb.2017.01.018

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135–168. doi:10.1146/annurev-psych-113011-143750

Ferrari, E., Robins, B., & Dautenhahn, K. (2009, September). Therapeutic and educational objectives in robot assisted play for children with autism. Paper Presented at the 18th IEEE International Symposium on Robot and Human Interactive Communication, Toyama, Japan.

Goldstein, J. (2012). Play in children’s development, health and well-being. Retrieved from https://www.ornes.nl/wp-content/uploads/2010/08/Play-in-children-s-development-health-and-well-being-feb-2012.pdf
González, M. R. (2015). Computational thinking test: Design guidelines and content validation. Paper presented at the EDULEARN15 Conference on Education and New Learning Technologies, Barcelona, Spain.

Hendry, A., Jones, E. J., & Charman, T. (2016). Executive function in the first three years of life: Precurors, predictors and patterns. Developmental Review, 42, 1–33. doi:10.1016/j.dr.2016.06.005

Highfield, K., Mulligan, J., & Hedberg, J. (2008). Early mathematics learning through exploration with programmable toys. In O. Figueras, J. L. Cortina, S. Alatorre, T. Rojano, & A. Sepúlveda (Eds.), Proceedings of the Joint Meeting of Psychology and Mathematics Education (Vol. 32, pp. 169–176). Morelia, México.

Hughes, C. (2011). Changes and challenges in 20 years of research into the development of executive functions. Infant and Child Development, 20(3), 251–271. doi:10.1002/icd.736

Janka, P. (2008). Using a programmable toy at preschool age: Why and how. Proceedings of the International Conference on Simulation, Modeling and Programming for Autonomous Robots (pp. 112–121), Venice, Italy.

Kazakoff, E. R., Sullivan, A., & Bers, M. U. (2013). The effect of a classroom-based intensive robotics and programming workshop on sequencing ability in early childhood. Early Childhood Education Journal, 41(4), 245–255. doi:10.1007/s10643-012-0554-5

Korkman, M., Kirk, U., & Kemp, S. (1998). Manual for the NEPSY. San Antonio, TX: Psychological Corporation.

Kort, W., Schittekatte, M., Compaan, E. L., Bosmans, M., Bleichrodt, N., Vermeir, G., ... Verhaeghe, P. (2002). WISC-III NL. Handleiding. London, UK: The Psychological Corporation.

Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). Neuropsychological assessment. New York, NY: Oxford University Press.

Mitnik, R., Nussbaum, M., & Soto, A. (2008). An autonomous educational mobile robot mediator. Autonomous Robots, 25(4), 367–382. doi:10.1007/s10514-008-9101-z

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howarter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. Cognitive Psychology, 41(1), 49–100. doi:10.1006/cogp.1999.0734

Phillips, L. H. (1999). The role of memory in the Tower of London task. Memory, 7(2), 209–231. doi:10.1080/741944066

Porteus, S. D. (1959). The maze test and clinical psychology. Oxford, England: Pacific Books.

Raven, J., Raven, J. C., & Court, J. H. (1998). Manual for Raven’s progressive matrices and vocabulary scales: Section 2: Coloured progressive matrices. San Antonio, TX: Pearson.

Regard, M., Strauss, E., & Knapp, P. (1982). Children’s production on verbal and non-verbal fluency tasks. Perceptual and Motor Skills, 55(3), 839–844. doi:10.2466/pms.1982.55.3.839

Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. Computers in Human Behavior, 72, 678–691. doi:10.1016/j.chb.2016.08.047

Strauss, E., Sherman, E. M., & Spreen, O. (2006). A compendium of neuropsychological tests: Administration, norms and commentary. New York, NY: Oxford University Press.

Sullivan, A., & Bers, M. U. (2016). Robotics in the early childhood classroom: Learning outcomes from an eight-week robotics curriculum in pre-kindergarten through second grade. International Journal of Technology and Design Education, 26(1), 3–20. doi:10.1007/s10798-015-9304-5
Sullivan, A., Kazakoff, E. R., & Bers, M. U. (2013). The wheels on the bot go round and round: Robotics curriculum in pre-kindergarten. *Journal of Information Technology Education: Innovations in Practice, 12*, 203–219. doi:10.28945/1887

Taylor, R. (1990). Interpretation of the correlation coefficient: A basic review. *Journal of Diagnostic Medical Sonography, 6*(1), 35–39. doi:10.1177/87564793900600106

Tucha, L., Aschenbrenner, S., Koerts, J., & Lange, K. W. (2012). The Five-Point Test: Reliability, validity and normative data for children and adults. *PLoS One, 7*(9), e46080. doi:10.1371/journal.pone.0046080

Unterrainer, J. M., Rahm, B., Kaller, C. P., Leonhart, R., Quiske, K., Hoppe-Seyler, K., ... Halsband, U. (2004). Planning abilities and the Tower of London: Is this task measuring a discrete cognitive function? *Journal of Clinical and Experimental Neuropsychology, 26*(6), 846–856. doi:10.1080/13803390490509574

Vander Steene, G., & Bos, A. (1997). *WPPSI-R: Wechsler preschool and primary scale of intelligence (Dutch-Flemish version)*. Lisse: Swets & Zeitlinger.

Vygotsky, L. S. (1967). Play and its role in the mental development of the child. *Soviet Psychology, 5*(3), 6–18. doi:10.2753/RPO1061-040505036

Welsh, M. C., Revilla, V., Strongin, D., & Kepler, M. (2000). Towers of Hanoi and London: Is the nonshared variance due to differences in task administration? *Perceptual and Motor Skills, 90*(2), 562–572. doi:10.2466/pms.2000.90.2.562

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35. doi:10.1145/1118178.1118215