



Dynamic testing: Can a robot as tutor be of help in assessing children's potential for learning?

Resing, W.C.M.; Bakker, M.; Elliott, J.G.; Vogelaar, B.

Citation

Resing, W. C. M., Bakker, M., Elliott, J. G., & Vogelaar, B. (2019). Dynamic testing: Can a robot as tutor be of help in assessing children's potential for learning? *Journal Of Computer Assisted Learning*, 35(4), 540-554. doi:10.1111/jcal.12358

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/85388>

Note: To cite this publication please use the final published version (if applicable).



Dynamic testing: Can a robot as tutor be of help in assessing children's potential for learning?

Wilma C.M. Resing¹  | Merel Bakker² | Julian G. Elliott³ | Bart Vogelaar¹

¹Developmental and Educational Psychology, Leiden University, Leiden, the Netherlands

²Instruction Psychology and Technology, Catholic University Leuven, Leuven, Belgium

³School of Education, Durham University, Durham, UK

Correspondence

Prof. dr. Wilma C. M. Resing, Faculty of Social Sciences, Department of Psychology, Section Developmental and Educational Psychology, Leiden University, PO Box 9555, 2300 RB Leiden, the Netherlands.
Email: resing@fsw.leidenuniv.nl

Abstract

This study examined whether computerized dynamic testing by utilizing a robot would lead to different patterns in children's (aged 6–9 years) potential for learning and strategy use when solving series-completion tasks. The robot, in a "Wizard of Oz" setting, provided instructions and prompts during dynamic testing. It was found that a dynamic training resulted in greater accuracy and more correctly placed pieces at the post-test than repeated testing only. Moreover, children who were dynamically trained appeared to use more heuristic strategies at the post-test than their peers who were not trained. In general, observations showed that children were excited to work with the robot. All in all, the study revealed that computerized dynamic testing by means of a robot has much potential in tapping into children's potential for learning and strategy use. The implications of using a robot in educational assessment were stressed further in the discussion.

KEYWORDS

computer, dynamic testing, educational assessment, inductive reasoning, robot, series completion

1 | INTRODUCTION

Recently, considerable development of new educational technologies, involving the use of seamless technology (Liu et al., 2014), tablets, and even robots, has triggered research into the effects of implementing these materials in educational settings (André et al., 2014; Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013). Recent research has focused on the use of robots in education, for example, as an instructional tool for transmitting knowledge (Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Chin, Hong, & Chen, 2014). In educational settings, robots can be classified based on how

they are used: as a tool (technology aid), peer, tutor, or a novice (Belpaeme et al., 2018; Mubin et al., 2013). Usage of (personalized) robot peers or tutors in educational learning and assessment procedures has gained attention in recent years (e.g., Baxter, Ashurst, Read, Kennedy, & Belpaeme, 2017; Belpaeme et al., 2013; Belpaeme et al., 2018; Benitti, 2012; Hong, Huang, Hsu, & Shen, 2016). One form of educational assessment where the use of robots may be particularly interesting is dynamic testing.

Whereas conventional, static test procedures are characterized by testing without providing the testee with any form of feedback, dynamic testing is based on the assumption that test outcomes resulting

We confirm that no substantial part of this manuscript has been previously published elsewhere or is under consideration for publication. We also confirm that there is no conflict of interest. There is no conflict of interest, and our research and handling of our data have been approved by the local Ethics Committee (CEP) of our university. All listed authors of this submission have seen and approved the manuscript.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Journal of Computer Assisted Learning* Published by John Wiley & Sons Ltd

from a scaffolded feedback procedure or intervention are more likely to provide a good indication of a person's level of cognitive functioning than conventional, static test scores. The primary aims of research in dynamic testing have been to examine progression in cognitive abilities following training between test session(s), to consider behaviour related to the individual's potential for learning, and to gain insight into learning processes at the moment they occur (Elliott, Grigorenko, & Resing, 2010; Resing, Touw, Veerbeek, & Elliott, 2017). Dynamic test procedures differ from static ones, because in a dynamic test situation testees are given (guided) instruction enabling them to show individual differences in progress when solving equivalent tasks.

The aim of the current study was to investigate whether a computerized one-on-one dynamic test administered by a tutor robot could allow for (investigating) systematic and controlled dynamic testing outcomes. In doing so, we sought to examine the effects of receiving instruction and training by a robot on children's changes in performance across test sessions.

A major difficulty in undertaking highly interactive forms of assessment is that the assessor must try to fully engage with the child while also recording in detail each step in the process. A key advantage of computerized testing is that it may be possible to register every task-solving step taken by the child, which would provide examiners with the opportunity to analyse the sequence of these steps. This would offer valuable information about the child's learning progression during the dynamic process (e.g., Resing & Elliott, 2011). Computerized assisted instruction provided by a personalized robot may also offer promising new possibilities for dynamic testing. These include using more flexible approaches to task-solving, using more adaptive scaffolding procedures, and, consequently, creating a more authentic assessment environment (Huang, Wu, Chu, & Hwang, 2008; Khandelwal, 2006). Therefore, the one-on-one tutor robot in the present study, which had an attractive appearance to children, was designed to detect the children's task-solving steps, provide hints to solve the tasks, record in detail children's responses to that assistance and react adaptively to children's solving behaviour.

1.1 | Dynamic testing of inductive reasoning

Conventional, static tests are often used by educational and school psychologists and are viewed as a satisfactory means of measuring previous learning. Dynamic test measures, on the other hand, often employing a test-training-test format, are designed to assess developing or yet-to-develop abilities (Elliott et al., 2010; Sternberg & Grigorenko, 2002). The theoretical framework for dynamic testing can be linked to the ideas of Vygotsky (1978), who posited that children's learning can be characterized as a social process, occurring in their zone of proximal development. This zone of proximal development has been defined in terms of the difference between children's independent task-solving, the *actual level of development*, and their level of task-solving after help or instruction has been given, often in the form of scaffolds, the *potential level of development*. The current study made use of robot-administrated structured pre-test and post-test instructions, and a *graduated prompts*

training procedure in between, consisting of two separate sessions in which children were provided with hints to help them solve the tasks. These prompts (or hints) included increasingly more specific and explicit feedback on how to solve the task presented. The hierarchical step-by-step provision of these *prompts* was given in accordance with the child's perceived needs based on their given solution of the task. In the current study, we programmed this hierarchical step-by-step procedure, in order to, potentially, examine the effectiveness of dynamic testing provided by a robot.

Dynamic testing studies often examine children's inductive reasoning ability (e.g., Bethge, Carlson, & Wiedl, 1982; Guthke & Beckmann, 2000; Hessels-Schlatter, 2002; Perret, 2015; Tzuriel, 2001; Stevenson, Heiser, & Resing, 2013; Tzuriel & Flor-Maduel, 2010; Vogelaar & Resing, 2016). Inductive reasoning tasks, for example categorization, inclusion, seriation or analogical reasoning, involve general rule finding processes, which require the detection of similarities and/or differences between task characteristics or in the relations between these characteristics under examination (Csapó, 1997; Klauer & Phye, 2008; Molnár, Greiff, & Csapó, 2013). Progression in task accuracy after training or repeated testing has been reported in a variety of inductive reasoning domains, but mostly in class-inclusion tasks (e.g., Siegler & Svetina, 2006), and matrices/analogies (Alexander, White, Haensly, & Crimmins-Jeanes, 1987; Alexander, Willson, White, & Fuqua, 1987; Passig, Tzuriel, & Eshel-Kedmi, 2016; Resing, Bakker, Pronk, & Elliott, 2017; Tzuriel & George, 2009; Vogelaar & Resing, 2016).

Our study focused on children's performance on series completion tasks, a subtype of inductive reasoning which has been shown to be a sensitive indicator of children's problem-solving ability (e.g., Holzman, Pellegrino, & Glaser, 1983; Molnár et al., 2013). We used a schematic-series completion task utilizing *puppets*: children were shown a series of puppets, with different arms, legs, bellies and heads; had to discover what the next puppet in the row had to look like; and had to construct the right puppet using tangible puzzle pieces (e.g., Resing & Elliott, 2011).

With regard to solving inductive reasoning tasks, a distinction has been made between analytical and heuristic strategies (e.g., Klauer & Phye, 2008). An analytical strategy requires investment of time in planning the solution, whereas a heuristic strategy utilizes more time to test and retest (partial) hypotheses about the solution process. For this reason, an analytical strategy is shown to require more time during the first, planning phase of solving tasks, whereas using a heuristic strategy requires more time for testing hypotheses about the solution. In the current study, we examined children's analytical and heuristic strategy use in solving series completion tasks.

Researchers have stressed the importance of using tangible objects in learning and assessment environments. Piaget (1955); Wood, Bruner and Ross (1976); and Vygotsky (1978) already argued for the use of concrete manipulatives in childhood education to develop mental representations and help children gain knowledge about the characteristics of these materials. Others have also emphasized the importance of using tangibles (Collins & Laski, 2015; Khandelwal, 2006; Khandelwal & Mazalek, 2007; Manches & O'Malley, 2016; Verhaegh, Fontijn, Aarts, & Resing, 2013; Wood & Wood, 1996). The series completion task in the current study employed three-

dimensional tangible puzzle pieces, which allowed the children to manipulate all the pieces freely, and enabled observation of their ways of solving the tasks.

1.2 | Computerized dynamic testing

To the authors' knowledge, no research has been conducted in which a robot was used to administer a dynamic test. Resing, Steijn, Xenidou-Dervou, Stevenson, and Elliott (2011), however, investigated whether computerized dynamic testing, using a multiple assessment, test-training-test format with graduated prompts given by a computer, provided more information about test performance than when these prompts were given by an examiner. Key to this graduated prompts approach is the possibility to incorporate feedback and tailored assistance into the training phases (Elliott et al., 2010; Grigorenko, 2009; Jeltova et al., 2011). Although no differences in accuracy were reported, the computerized version of the dynamic test provided more detailed information on the individual task-solving processes. In an earlier study, Tzuriel and Shamir (2002) reported that children who were assisted by a computer when taking the Children's Seriation Thinking Modifiability Test showed more cognitive change than when feedback was provided by an examiner. Such findings, in combination with the seamless learning possibilities a robot offers, and the tangible 3D-task used in this study, suggest that computerized assisted instruction provided by a robot may offer promising new possibilities for dynamic testing.

In the current study, a small, friendly tutor robot was utilised during a sequence of test sessions. We developed a Wizard of Oz setting (Dahlbäck, Jönsson, & Ahrenberg, 1993), with the examiner partially operating the robot by computer. The behaviour of the robot was preprogrammed to be adaptive to the child's responses and incorrect task-solving behaviour, based on the outcomes of former studies (e.g., Resing et al., 2017). Furthermore, the robot was preprogrammed to provide oral prompts and scaffolds (individual hints based on the children's actions) when solving the inductive reasoning tasks. In addition, it was programmed to give general feedback and short instructions and interact nonverbally by, for instance, naming the child, nodding, dancing and blinking its eyes. The robot was tele-operated, so it had to be controlled by an examiner. The current study sought to investigate the potential of using this robot as an assessment tool for children when solving reasoning tasks in a dynamic testing context. We also aimed to get a first impression of the interactions between children and the robot for the development of an optimal and authentic learning and assessment environment.

1.3 | Use of robots in education

The use of robots in education, for example, for psycho-educational testing and assessment, has been associated with several advantages. Robots are well suited to physically and socially engage with learners and their environment, with learners showing more social behaviour beneficial for learning and increased learning gains vis-à-vis other forms of technical support that do not have a physical embodiment (Belpaeme

et al., 2018). Positive effects of the use of robots compared with other forms of technical assistance have been found in the cognitive as well as the affective domain. With regard to the affective domain, recent studies in the field of education have examined the use of technology to support and increase children's motivation in classrooms (Chin et al., 2014). The presence of robots has been found to have a positive influence on children's motivation when solving cognitive tasks (André et al., 2014). A robot's movement and body gestures appear to be interesting motivators that could affect a respondent's decision-making processes (Shinozawa, Naya, Yamato, & Kogure, 2005). Furthermore, studies have shown that robots designed to express social cues positively influenced respondents' motivation to finish a task and increased their desire to spend more time with the robot (Tanaka, Cicourel, & Movellan, 2007). In various studies, robot characteristics such as appearance, mobility and animation have been shown to influence even kindergarten children's ability to learn from robotic instructions and sustain their interest in completing tasks (e.g., Brown & Howard, 2013).

In relation to the cognitive effects of robots, young children have shown their ability to learn from a peer or tutor robot in several domains, such as vocabulary performance, (second) language learning, mathematics, science, thinking skills and self-regulated learning (Chang, Lee, Chao, Wang, & Chen, 2010; Hussain, Lindh, & Shukur, 2006; Jones & Castellano, 2018; Moriguchi, Kanda, Ishiguro, Shimada, & Itakura, 2011; Movellan, Eckhardt, Virnes, & Rodriguez, 2009; Sullivan, 2008). In such studies, participants demonstrated positive and engaging interactions with the robot. André et al. (2014) showed that robots could influence children's behaviour positively when they were given mental arithmetic tasks. Several authors have reported that robot-based instruction methods could have similar effectiveness as human instructors (Brown & Howard, 2013). Reaching a similar conclusion, Serholt, Basedow, Barendregt and Obaid (2014) noted that the children in their study asked the human instructor more often for help. They also concluded that children were able to follow instructions from a robot but added that more long-term interaction between subjects and a robot would be needed for studying lasting effects. In their overview of studies in the field of early language learning, however, Kanero et al. (2018) concluded that social robots are useful in language learning but not (yet) as effective as human teachers.

The current study examined the use of a robot during multiple assessment sessions. We sought to examine the effects of receiving instruction and training by a robot on children's performance across these test sessions.

1.4 | Current study aims

In the light of the promising findings about children's engagement with robots in the classroom (e.g., André et al., 2014; Baxter et al., 2017; Belpaeme et al., 2018; Benitti, 2012; Deublein et al., 2018; Kozima & Nakagawa, 2007; Tanaka et al., 2007), we sought to examine the potential of utilizing a tutor robot in a dynamic testing setting, as a means to interact with the children and to record their performance. We focused on four key underlying issues.

Our first task was to examine the effect of training with graduated prompts, provided by the robot, on children's inductive reasoning performance. We expected that trained children would demonstrate larger increases in pre-test-post-test progressions in accuracy and the number of correctly solved pieces of the series completion task compared with untrained children. These expectations were in accordance with findings from Resing, Xenidou-Dervou, Steijn, and Elliott (2012), Stevenson, Touw, and Resing (2011), Tzuriel and Shamir (2002), Passig et al. (2016) and Wu, Kuo and Wang (2017).

Secondly, we investigated children's need for instructions, provided by the robot, during training, and expected that the number of prompts children required during the training sessions would decrease from training 1 to training 2, indicating a learning effect (Authors, 2011). In doing so, we inspected the types of prompts provided separately (metacognitive, cognitive and modelling; Resing & Elliott, 2011).

Thirdly, we examined whether training would influence children's strategy use by examining how their inductive reasoning performance changed at a behavioural level. We expected a change towards a more advanced, analytical strategy level for trained children only (Resing et al., 2012).

In addition, we explored individual differences in the progression of children as a consequence of dynamic testing. The trained children were split into groups on the basis of the number of prompts they needed during training, in combination with lower or higher pre-test scores. We explored whether the progression paths in inductive reasoning of the various groups of children were significantly different (Resing et al., 2017).

2 | METHOD

2.1 | Participants

Fifty-two 8-year-old children with a mean age of 96 months ($SD = 7.2$ months; range = 83–116 months) participated in this study. The children, 26 girls and 26 boys, were recruited from four second and third grade classes of middle-class elementary schools, located in the western part of the Netherlands. All children were born in the Netherlands, and Dutch was the first language spoken at school and at home. The schools were selected on the basis of their willingness to participate. Prior to the study, written informed consent was obtained from the schools and parents. The testing was undertaken by three trained postgraduate students with teaching experience.

TABLE 1 Study design

Condition	Group ^a Raven	Session 1 Pre-test	Session 2 Training 1	Session 3 Training 2	Session 4 Post-test
Training	X	X-R	X-R	X-R	X-R
Control	X	X-R	Alternative tasks	Alternative tasks	X-R

Note. R: the robot was available on the child's desk.

^aThe Raven's progressive matrix test was administered in class, before dynamic testing started.

One child was not present during the administration of the second training session; his data were not included in any of the analyses. This research project was approved by the ethics board of our university.

2.2 | Design

The study employed a pre-test-training-post-test control-group design (see Table 1) with randomized blocking on the basis of children's scores on Raven's progressive matrices (Raven, Raven, & Court, 2003) administered before dynamic testing started. This blocking procedure is often used in studies with rather small experimental and control groups to assure that both groups do not differ very much with regard to an important variable of study, in this particular case the mean level of reasoning. Children's scores of the Raven test were ordered from high to low, and pairs were made of children with equal scores, etc. On the basis of this blocking procedure, administered per school and grade, children were, per pair, randomly assigned to either a *dynamic test group* (training condition: pre-test, training, and post-test) or a *static control group* (control condition: pre-test, control task, and post-test). In each school and grade, 50% of the children were allocated to the training condition; the others were assigned to the control condition. Children in both conditions were administered the pre-test and post-test of the series completion task (Sessions 1 and 4; see Table 1), with the robot providing the instructions. Children in the training condition were administered a short training in between (two times: Sessions 2 and 3), whereas in the same time window, control-group children completed other cognitive tasks, such as mazes and dots-to-dots tasks (Sessions 2 and 3). The robot was present during all sessions. The pre-test and post-test tasks took approximately 30 min to administer, and the two training sessions took about 20 min each.

3 | MATERIALS

3.1 | Raven's progressive matrices

The Raven's progressive matrix test (Raven et al., 2003) measures the ability to detect rules by means of induction, a prerequisite for successful inductive/serial reasoning. Each item is composed of a visual-spatial 3×3 matrix in which one part is missing. Children were instructed to select the missing piece from a number of alternatives. Split-half coefficients were reported as measure of the reliability of the test ($r = 0.91$; Raven, Raven, & Court, 2000).

3.2 | The robot

In this study, a small table-top robot, developed by WittyWorX (2012), was utilized. The robot had an appearance similar to a wise but friendly owl (see Figure 1). It was about 20 cm tall, and could easily be placed on a child's desk. The robot was preprogrammed to speak, dance, move, show feedback with its eyes, and react to touch. Non-verbal behaviour included emotions (happy/neutral) as shown by the eyes (two colour displays), nodding or head-shaking and dancing (body movement was possible in all directions). With its sensors and expression abilities, it was expected that the robot could interact with the children playfully and hold their attention.

The robot's stand-alone abilities were not fully developed at the time of testing. As mentioned, we utilized a Wizard of Oz setting, and the examiner, quietly sitting in a corner of the room behind the child, served as the eyes and ears of the robot. For sensory input, the robot was equipped with a camera, microphones and touch sensors, so that the solving processes of the children could be filmed (only capturing hands and voice to safeguard children's anonymity). All robot behaviour (and that of the examiner) was preprogrammed by using if-then scenarios with utterances, sounds and eye/body movements. As the examiner functioned as the eyes and ears of the robot and could follow the filmed solving behaviour of the child on a laptop, she could, by pushing a button on this screen, influence the preprogrammed behaviour path of the robot, according to the fixed scenarios. The robot was programmed in such a way that it was able to interact and give feedback at the right time. Children could press the head of the robot to indicate that they were finished with a task and ready for the next one.

3.3 | Dynamic test: Series completion

We used a dynamic visual-spatial task adapted from Resing and Elliott (2011). During the pre-test and post-test sessions, children were given series of schematic-picture completion problems that all consisted of a line of six puppet pictures printed in a booklet, followed by an empty box with a question mark (see Figure 2). They were asked to construct the seventh puppet on a white empty frame on their desk by placing eight transparent perspex pieces into the right configuration. Items could be solved by observing the systematic changes that occurred in the row and uncovering the underlying solution rule(s) (Resing & Elliott, 2011). Each answer had to consist of one head, two arms, two legs and a torso comprised of three pieces. Item difficulty level was dependent on the number of transformations in the row and the frequency of recurring patterns (periodicity). For each new series problem, children had to find new rules because they entail new, unknown strings of recurrently repeating elements (solving rule 1) and unknown changes in the relationship between these elements (transformations; solving rule 2). The pre-test and post-test both consisted of 12 incomplete puppet items and were constructed as parallel versions with equivalent items that could consist of the following transformations: gender (male, female), pattern (stripes, dots or plain), or colour (yellow, pink, green and blue). The pre-test and post-test booklets were systematically changed and differed only slightly, for example, by changing female puppets into male puppets or changing colours. In a recent study with the same materials (in a tangible format) internal consistencies for the pre-test ($\alpha = 0.74$) and post-test ($\alpha = 0.78$), and a test-retest reliability ($r = 0.78$) were reported (Veerbeek, Vogelaar, Verhaegh, & Resing, 2019).

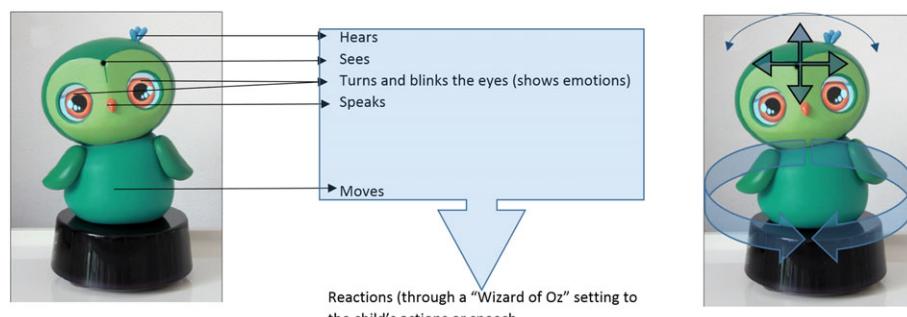


FIGURE 1 The robot owl-robot used in the current study [Colour figure can be viewed at wileyonlinelibrary.com]

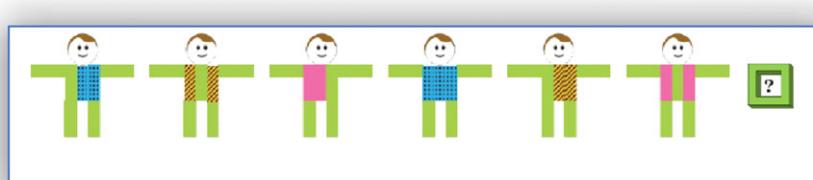


FIGURE 2 Example item of the series completion task [Colour figure can be viewed at wileyonlinelibrary.com]

At the start of both test sessions, the robot gave a general introduction to the children, explaining the purpose of the task. The children then had to place the pieces of the puppet in a frame and were required to remove them when the robot said so. The robot gave all the necessary oral instructions (through the preprogrammed voice files), and, if logically necessary, warnings (e.g., "First remove all the pieces from the plate" and "Keep the pieces within the lines"). Appendix A, as an example of the robot-child interactions, presents schematic overview of the Task Introduction as provided by the interaction between robot and examiner; the first two columns describe what the examiner (T. E.) had to say or do (press button); the next two columns give information on the preprogrammed voice samples; the last two columns provide information on the preprogrammed sounds.

The children were then given a practice item before starting with the various puppet items, following comparable preprogrammed scenarios. After this, the children started solving the task on their own. After constructing the solution, children had to touch the robot's head and were then asked, by the robot, to explain why their answer was correct ("Tell me, why does your puppet belong in the empty box?"). They received only marginal feedback from the robot at the introduction phase, like "well done" if a puppet item was solved correctly and "well tried" if time was up or a puppet item was solved incorrectly. Pre-test and post-test were further administered without any feedback.

3.4 | Dynamic test: Training

The 2×6 items used during the two training sessions were constructed at the same difficulty level and equivalent to those used during the pre-test and post-test sessions. However, children were now told at the start of the training sessions that the robot would help them find the correct puppet. The training procedure was based on graduated prompt procedures adapted from previous studies. This training procedure was developed on the basis of earlier developed

process models of the specific dynamic test utilized in the current study (e.g., Resing & Elliott, 2011; Resing et al., 2017). This so called *graduated prompts* procedure provided children with prompts to help them to solve the problem. These prompts included increasingly more specific and explicit feedback of how to solve the task presented. The hierarchical step-by-step provision of these prompts was given in accordance with the child's perceived needs based on their given incorrect solution of the task. Children were provided with stepwise instructions starting with general, metacognitive prompts such as focusing attention. Cognitive scaffolds were then offered, such as showing the pattern of the clothes of the puppets, and the robot could explain, if necessary, the underlying nature of the changes. When these prompts did not help the child to construct the correct puppet, the robot gave more individualized step-by-step guidance. Here, for example, the children were instructed by the robot (by means of the examiner who pushed a button) to get the correct body part of the puppet and put it in the box. During this part of the training, the children were given guidance for each body part of the puppet, if necessary. Two different sounds were used to indicate whether or not the body part the children had chosen was correct. If their puzzle piece was incorrect, the robot provided the correct one. After an item was solved in the frame on the table, children had to touch the robot's head. Then, the children were asked by the robot to explain why the puppet was correct ("Tell me, why does your puppet belong in the empty box?"). A schematic overview of the training procedure can be found in Figure 3. The robot was programmed to provide both oral and behavioural feedback. The robot shook its head and had blinking eyes when their answer was incorrect or nodded and said "Well done!" (with different blinking eyes) when the answer was correct.

3.5 | Procedure

The four test sessions took place once a week. All children were seen individually in their school during all four sessions. During the training, the robot interacted with the child and gave feedback and prompts

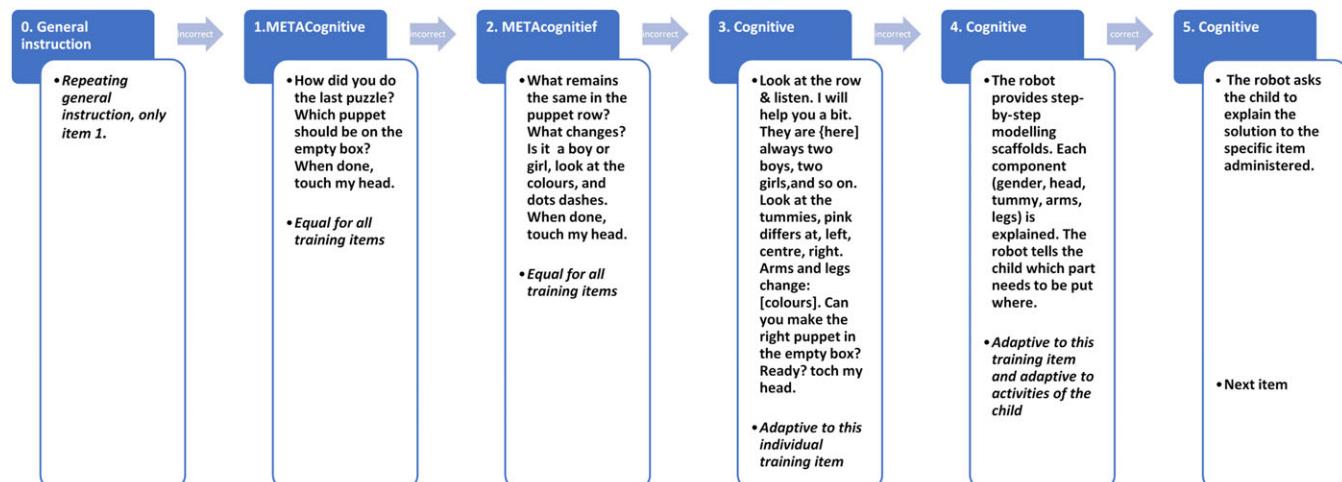


FIGURE 3 Schematic overview of graduated prompts training [Colour figure can be viewed at wileyonlinelibrary.com]

following the graduated prompts protocol. After the completion of each item, the child was asked to touch the head of the robot. For each item, the robot asked the children to explain why the puppet they had constructed was correct. The answers of all children were audio-recorded and videotaped, capturing only their hands to safeguard their anonymity, yet showing how they manipulated the materials during the test and training sessions. The camera was used to film the behaviour of the child, thereby allowing us to register time, taken position of the pieces and order of movement. Video-materials were only used to check the quantitative data. Every step that was taken by child the (and the robot) was saved in a log file.

The robot performed all the interactions with the child during the four sessions, operating with the help of the examiner who sat silently in a corner of the room, being part of our "Wizard of Oz" constellation. Voices and sounds uttered by the robot were actually initiated by the examiner who followed the task-solving behaviour of the child on the computer screen and had to push a button before the robot could execute the next step. The camera enabled the examiner to simultaneously analyse in detail the task-solving behaviour of children during the dynamic test sessions. The series completion items of all parts of the dynamic test were simulated on a laptop, and the examiner had to mimic the task exactly and at the very same moment as the child.¹ The robot was programmed in such a way that it was able to interact and give feedback at the right time.

3.6 | Scoring

3.6.1 | Number of prompts

We counted the number of times children received a prompt at least once for each item. The maximum number of prompts was 30 as there were six items and five types of prompts per item.

3.6.2 | Learner groups

The trained group of children was split into four learner groups: two groups of children that needed many versus few prompts during training, differentiated by those who had low or high pre-test scores. Median splits were used to separate the children into the four groups.

3.6.3 | Behavioural strategy use

The data gathered during the dynamic test sessions were compiled into log files. The outcome variables analysed were related to accuracy, time, efficiency and task-solving behaviour. Scoring of children's behavioural strategies during the test sessions was based on the observed solution times (ST) at different stages in the task-solving process (Kossowska & Nęcka. 1994): the initial period, which referred to the time before the first body part was placed; the middle ST, which referred to the period before the next piece was placed; and, lastly, the end ST, which referred to the total time it took children to solve the problem.

$$\text{Behavioral Strategy Use} = \frac{\text{InitialST} + \text{MiddleST}}{\text{InitialST} + \text{MiddleST} + \text{EndST}} * 100.$$

Higher scores on the Behavioural Strategy Use measure were thought to reflect the use of an analytical strategy as children spent relatively more time on the preparatory stage (initial and middle ST) of task-solving. Lower scores were assumed to reflect a heuristic strategy, indicating that the children took more time for the execution stage than the initial and middle stages. Children with low scores were likely to have thought more globally about what the last puppet should look like (Resing et al., 2012).

4 | RESULTS

Two one-way analyses of variance (ANOVAs) were conducted to examine whether there were any differences between the two treatment groups regarding their initial level of inductive reasoning and age. The analyses revealed that the two treatment groups did not significantly differ with regard to their average age ($F[1, 49] = 3.47, p = 0.07$) nor their initial reasoning performance at pre-test ($F[1, 49] = 0.74, p = 0.39$).

4.1 | Effects of training

First, children's performance on the series completion task and the effect of receiving training by the robot on their reasoning progression was analysed, regarding two outcome variables: (1) accuracy, measured as the total number of correctly constructed puppets at pre-test and post-test, and (2) total number of body parts positioned correctly at pre-test and post-test.

4.1.1 | Total correct

We expected that the trained children would show greater progression in reasoning accuracy than the children in the control group. The effect of training on accuracy was examined using a repeated measures ANOVA with Condition (training/control) as between-subjects factor and Session (pre-test/post-test) as within-subjects factor. The number of accurately solved items was the dependent variable. The change in reasoning accuracy across sessions is depicted in Table 2 and Figure 4. A significant effect of Session was found (Wilks' $\lambda = 0.92, F[1,49] = 4.22, p = 0.045, \eta_p^2 = 0.079$). More importantly, both groups of children showed different progression paths from pre-test to post-test, which was indicated by a significant interaction effect of Session and Condition (Wilks' $\lambda = 0.83, F[1,49] = 9.81, p = 0.003, \eta_p^2 = 0.167$). Consistent with our hypothesis, children in the training group showed significantly more progression in accuracy from pre-test to post-test than those in the control group. Large individual differences in progression were found as well, as the ranges of progression in the number of correct items show (experimental condition: from -2 to +6, range = 8; control condition: from -3 to +3, range = 6).

TABLE 2 Mean scores and standard deviations on accuracy and number of body parts at pre-test and post-test, per condition

Condition	Pre-test Mean (SD)	Post-test Mean (SD)
Total correct		
Control	4.56 (2.73)	4.24 (2.67)
Training	5.19 (2.51)	6.73 (2.93)
Total	4.88 (2.61)	5.51 (3.05)
Number of body parts		
Control	74.36 (11.39)	71.88 (14.15)
Training	74.77 (13.86)	79.77 (14.71)
Total	74.57 (12.58)	75.90 (14.57)

4.1.2 | Total body parts correct

The number of body parts children had positioned correctly at pre-test and post-test was analysed with a repeated measures ANOVA with Condition (training/control) as between-subjects factor and Session (pre-test/post-test) as within-subjects factor. The progression in the number of correct body parts for children in both conditions is depicted in Figure 4 (and Table 2). A nonsignificant effect of Session was found (Wilks' $\lambda = 0.98$, $F[1,49] = 1.07$, $p = 0.306$, $\eta_p^2 = 0.021$). Children in both treatment conditions did, however, progress differently from pre-test to post-test, which was indicated by a significant interaction effect of Session and Condition (Wilks' $\lambda = 0.84$, $F[1,49] = 9.41$, $p = 0.004$, $\eta_p^2 = 0.161$). Again, as expected, the training group showed a greater progression in the number of correctly positioned body parts from pre-test to post-test than the control group children who showed no progression. Again, large differences in progression of the number of body parts correctly placed were visible (experimental condition: from -10 to 20, range = 30; control condition: from -36 to 13, range = 49).

4.2 | Prompts during training

The number of prompts children needed during training was considered to be one of the indicators of their potential for learning. Children showed large individual differences in the number of metacognitive (training 1: ranging from 0 to 11; training 2: from 0 to 12) and cognitive (training 1: from 0 to 15; training 2: from 0 to 18) prompts. Contrary to our expectations, this did not significantly decrease from training 1 to training 2 ($t[25] = 0.35$, $p = 0.78$) as has been depicted in Figure 5. Further analyses after dividing the total number of prompts into the three types of prompts provided (metacognitive, cognitive and modelling) did not give different results ($p > 0.05$). Data also showed that children needed a considerable number of metacognitive prompts. The combination of data indicates that although the dynamically trained children still needed a number of prompts, they showed clear progress in their reasoning accuracy.

4.3 | Completion time and behavioural strategy use

The effect of training on children's total completion time and the time they needed for different task-solving stages at pre-test and post-test was also examined. First, a repeated measures ANOVA with Condition (training and control) as a between-subjects factor, Session (pre-test and post-test) as a within-subjects factor, and the total completion time as dependent variable. The analysis revealed a significant effect of Session (Wilks' $\lambda = 0.803$, $F[1,49] = 12.00$, $p = 0.001$, $\eta_p^2 = 0.297$) and a nonsignificant Condition \times Session effect (Wilks' $\lambda = 0.95$, $F[1,49] = 2.76$, $p = 0.103$, $\eta_p^2 = 0.053$). On average, the children in both the training and the control condition decreased their total completion time from pre-test to post-test (see Table 2).

We then investigated whether the behavioural strategy use of the children in the training and control conditions changed differently from pre-test to post-test. Another repeated measures ANOVA was conducted with Session (pre-test and post-test) as a within-subjects

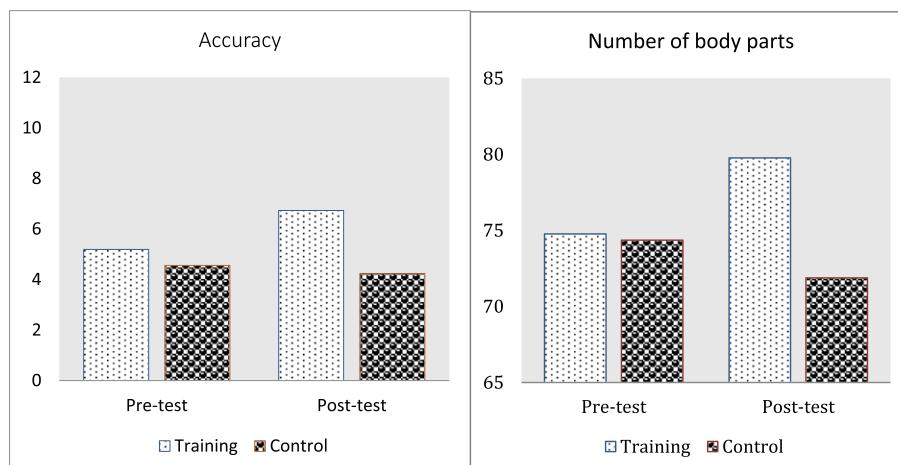


FIGURE 4 Progress in accuracy, and number of correctly positioned body parts for trained and control group children across session (pre-test/post-test) [Colour figure can be viewed at wileyonlinelibrary.com]

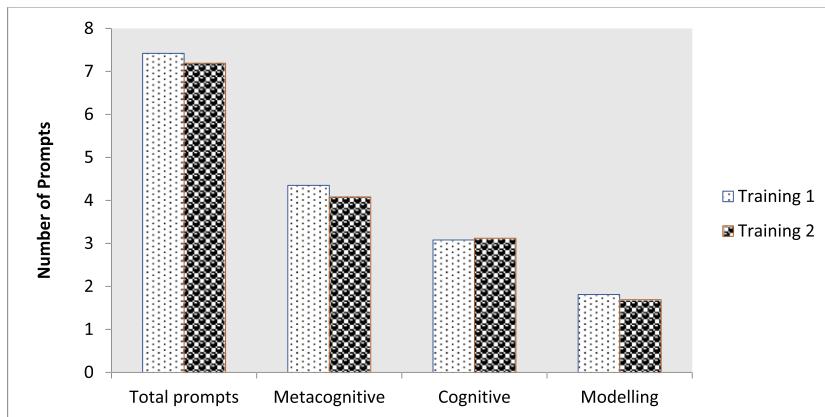


FIGURE 5 The total prompts, and the metacognitive, cognitive and modelling prompts [Colour figure can be viewed at wileyonlinelibrary.com]

factor, Condition (training and control) as a between-subjects factor, and Behavioural strategy as the dependent variable. A non-significant Session effect was found (Wilks' $\lambda = 0.993$, $F[1,49] = 0.36$, $p = 0.549$, $\eta_p^2 = 0.007$). Figure 6 shows that, on average, children's relative preparation time did not change. The significant interaction effect between Session and Condition (Wilks' $\lambda = 0.880$, $F[1,49] = 6.70$, $p = 0.013$, $\eta_p^2 = 0.120$), however, indicates that the dynamic training differentially influenced children's behavioural strategy use. The children in the training condition scored significantly lower on the post-test, thereby, unexpectedly, making more use of heuristic strategies after training, whereas the control group children appeared to use the more analytical strategies more frequently at the post-test.

4.4 | Exploring learner groups

In addition, we explored individual differences in the progression of children. The trained children were split into four learner groups: needing many versus few prompts during training, in combination with lower or higher pre-test scores. The low pre-test and low prompts group included only two children and was not included in the analysis. Figure 7 reveals the progression paths of the children in the three groups. A repeated measures ANOVA, with the number of accurately solved items as dependent variable, Session as a within factor and

Pre-test-Prompts category as a between factor, revealed a significant effect of Session (Wilks' $\lambda = 0.733$, $F[1,21] = 7.63$, $p = 0.012$, $\eta_p^2 = 0.267$) and a nonsignificant Pre-test-Prompts category \times Session effect (Wilks' $\lambda = 0.95$, $F[1,49] = 2.76$, $p = 0.103$, $\eta_p^2 = 0.053$). On average, all groups increased their performance from pre-test to post-test, although, there was no significant difference in the rate of progression for the three groups, although a trend in the expected direction can be seen ($p = 0.10$ for the interaction effect). At an individual level, in all groups some children showed large progression, and others did not, indicating that even in the groups of children starting at a lower pre-test level, some children profited considerably of the training, whereas others did not show such progression.

4.5 | Observations

From the outset of this study, children were highly excited and motivated to work with the robot, which appeared to know every child by name. They liked the testing periods very much and were eager to work with the robot. After a short period of time, they were talking to Myro as if it was a teaching assistant, and most of the time ignored the examiner who was sitting in a corner of the room. Because the instructions provided by the robot were highly structured, they sometimes pushed it on the head and said things like: "keep your



FIGURE 6 Behavioural strategy-scores at pre-test and post-test for training and control group children

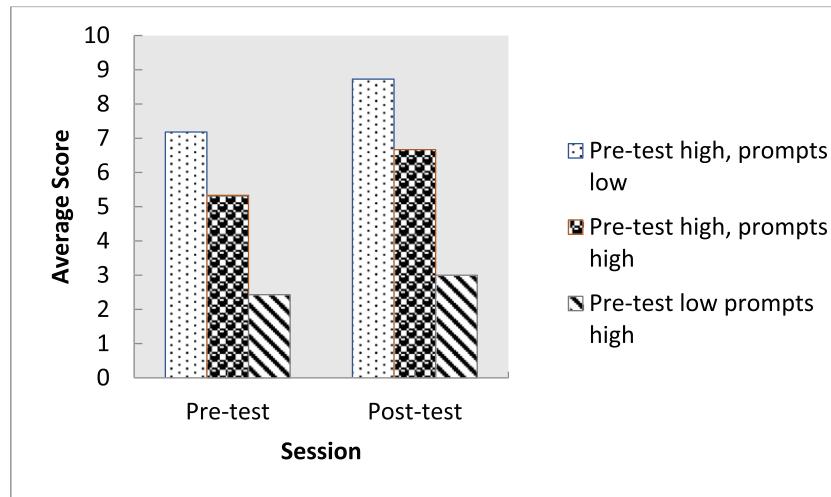


FIGURE 7 Change in progression from pre-test to post-test, per learner group [Colour figure can be viewed at wileyonlinelibrary.com]

mouth shut; you have said that now too often, Myro." Their teachers also responded enthusiastically, many asking if they could play the game with the robot, so a general meeting was planned after the study ended.

5 | DISCUSSION

The present study focused on the potential of using a pre-programmed table-top robot in a Wizard of Oz setting as an educational assistant and training tool for primary school children. In line with previous studies on children's seriation and analogical reasoning skills (e.g., Freund & Holling, 2011; Resing & Elliott, 2011; Stevenson et al., 2013), our study showed that task performance generally improved when children were tested twice, but that the degree of progression varied, depending on whether or not children were dynamically trained by the robot on the task (e.g., Campione & Brown, 1987; Passig et al., 2016; Resing et al., 2011, 2017, 2012). Children that were dynamically tested and trained by the robot showed significantly greater progression in both their accuracy of task solving and the more detailed *number of correct puzzle pieces* variable than children who were just statically tested by the robot. We believe that we can safely conclude that the intervention children were provided with by our friendly table-top robot led to these differences in progression because the same tasks and instructions were tested and positively evaluated in other studies (e.g., Resing & Elliott, 2011; Resing et al., 2017; Veerbeek et al., 2019). Of course, a future study, in which a second control group solves the items used in both training sessions in a static, unguided way, would provide extra information, necessary to further confirm this conclusion. Dynamic testing research in the past with children in these three conditions already provides further support to this conclusion (e.g., Resing, 1993, 2000). Another useful direction for future studies concerns investigating the potential advantages of robot-administered as opposed to computerized or human-administered dynamic tests to research whether robot-administered dynamic testing has benefits beyond those of human and computerized testing.

Interestingly, children's progression paths increased, whereas the number of prompts they needed did not decrease from the first to the second training. This could be partially due to the difficulty level of the series-completion items; they were developed as rather difficult tasks on purpose. More or more enduring training periods could, possibly, result in children showing extra progression in task solving as well as then needing fewer prompts or scaffolds while being trained. Children showed large individual differences in the number of prompts they needed. Further research with a revised design might provide more information in the future. Another potential reason might be that the children experienced the robot as such a nice companion that they wanted to continue receiving prompts and scaffolds from it.

The scaffolding and graduated prompts principles behind the training given by the robot were specifically designed to tap into children's zone of proximal development (Serholt & Barendregt, 2016; Vygotsky, 1978). When we explored the variation in progression in task solving in relation to the outcomes, large individual differences were detected. Of course, the data regarding learner groups are rather speculative but, considering the small subgroups of children, are promising, and highlight the potential extra value of individualized forms of dynamic testing, in particular with computerized robot technology. In future, outcomes of an extended study will have to support these preliminary findings.

The current study shows that our dynamic training provided by a robot did also differentially influence children's behavioural strategy use as measured by the time children needed to actually start solving each task item. Unexpectedly, the trained children made less use of a more analytical strategy after training than their peers who did not receive training. The untrained children, however, appeared to use an analytical strategy more frequently during the post-test. Nevertheless, our first, global checking of the log files revealed that trained children more systematically placed the puppet blocks; they first selected little piles of equal blocks, for example, three green ones for the body of the puppet; then made a three-piece block of the body, and finally placed that 3×3 block on the puppet frame. Untrained children, on the contrary, frequently seemed to use quick trial-and-error behaviour

or solved the puzzle piece by piece. Perhaps the unexpected findings regarding the increase in heuristic strategy use by the trained children reflects familiarity with the task, as a result of which these children required less preparation. This finding underlines that we cannot solely rely on reaction time data in relation to children's behavioural strategy use (e.g., Kossowska & Nęcka, 1994), but, of course, future research with a larger sample size will be necessary to underline our findings and inferences with regard to children's strategy use. Step-by-step analysis of children's task-solving sequences would be one possible option (e.g., Resing et al., 2017; Veerbeek et al., 2019).

The results further support our idea that subgroups can be discerned that differ on the basis of their changing strategy use, particularly in the case of the trained children, in combination with information regarding the number of prompts children need during training, and their progression in accuracy and strategy use. Findings lend further support to the idea that dynamic testing outcomes can be helpful for educational assessors because these provide interesting process information regarding inter-variability and intra-variability in children's use of strategies when learning to solve tasks.

In the current study, the robot provided prompts to the child when needed, but these were not yet optimally adaptively tailored to the, at times, very idiosyncratic mistakes that the children incidentally made during training. Further research is necessary to ensure that the robots of tomorrow provide highly sophisticated and differentiated interaction responses in assessment contexts. With regard to the cognitive domain studied here, future research should be geared to the fine-tuning of prompts, and dynamic scaffolds, adaptations to specific groups of children, examination of specific, systematic task-analyses, and consideration of patterns of mistakes and idiosyncratic ways of processing children show in solving cognitive tasks (e.g., Granott, 2005; Khandelwal, 2006; Renninger & Granott, 2005). In future, for example, the robot could be programmed to enable more variation and flexibility in preprogrammed scenarios in providing feedback and instruction to individual children. Although we are aware it is a challenge to realize all these requirements, these developments should provide exciting possibilities for obtaining further insight into children's differing learning paths during dynamic testing, or in relation to instruction in the classroom. Although the robot still had some obvious limitations, such as repeating instructions in exactly the same way, and the robot was operated in a "Wizard of Oz" setting, children interacted with the robot freely, for example, providing the robot with feedback, and were highly responsive and motivated to work with the robot, even after all the assessment and training sessions. The vast majority of children did not even seem to notice that the examiner was seated in the back of the room.

A particular complication of dynamic testing, in particular when individual strategy patterns and changes are the focus of assessment, is that detailed study of children's processing, including their responses to training, can easily result in an overload of information (derived from spoken, written or videotaped sources) that is too complex and time-consuming to interpret and report. A personalized robot teacher assistant would certainly help to overcome this difficulty, especially if it would be able to visually deal with the pieces of tangible

materials children put on the table freely. We think this is a key and unique aspect of using robotics in psychological and educational assessment, because both the development and education of higher cognitive abilities have their origin in sensory-motor activities in young children (e.g., Timms, 2016), and the robot in combination with the material as developed perfectly match these activities. We anticipated and found that such technology can assist us in assessing and examining task-solving processes in more detail, thereby enabling us to inspect in-depth more of the information processing that takes place during the course of training, one of the key elements of process-oriented dynamic testing (Elliott, Grigorenko, & Resing, 2010; Jeltova et al., 2011; Sternberg & Grigorenko, 2002). As most empirical studies that discuss the effects of robots as teaching tools involve learning closely related to the field of robotics, our findings have significant potential and should provide further opportunities for the broader field of learning complex reasoning skills (Benitti, 2012).

We are aware that much effort in terms of both hardware and software development will be necessary for educational assessment before educational robots will be ready to assist teachers and educational psychologists in the classroom of tomorrow (e.g., Timms, 2016). We think, however, that the results of the current study reveal that even a simplified version of a real robot, as a result of its instructive teaching and patience, can stimulate children in their learning of solving complex reasoning tasks, leading to an important impact on the development of cognitive growth (Mubin et al., 2013). We noted that the children enjoyed the testing periods very much and were eager to work with the robot during all assessment sessions. It would be valuable to study whether children in the *control* condition also learned a lot from assessment by the robot, as they also were eager to leave the classroom for a next session with the robot. An extension of the study design with a focus on the novelty aspect of a robot-administered dynamic test will therefore be necessary in future, investigating the effect of repeated interactions of the robot as possible influences on the outcomes. Possible examples of such influences include being distracted by the robot or the magnitude of the cognitive load posed by robot-administered tasks. The focus of the current study was on quantitative analysis of children's (cognitive) changes brought about by being assessed by a robot. Future studies could focus more on qualitative analyses, for instance, by analysing qualitative differences in children's approach to solving tasks.

The merits of using a robot as an assistant in dynamic testing are, of course, intriguing. Earlier studies (Resing & Elliott, 2011; Resing et al., 2017) have already highlighted the benefits of the use of an electronic console for dynamic testing. Our study replicated the potential of electronic technology for dynamic testing but also introduced the robot as a helpful coassessor, whereby the children could freely play with the tangibles, organizing and moving them. The robot has been found to be an enjoyable dynamic companion, mostly because it possessed both verbal and nonverbal interaction qualities, with—for the moment—the examiner as Wizard of Oz at the background. Earlier research also showed that the use of a preprogrammed

computerized interface for offering the prompts and scaffolds has no discernible negative consequences when compared with that provided by an examiner (Stevenson et al., 2011; Tzuriel & Shamir, 2002). Because the task prompts and scaffolds remained the same over studies, we think that these earlier findings are generalizable to the outcomes of the current study. Nevertheless, it will be necessary to check the potential application of an assistant-robot assessor in comparison to both a human and a (2D) computer administration of the dynamic test, to further validate the additional value of robot-administered dynamic testing. Our recommendations for future studies would be to continue to explore possibilities in the use of preprogrammed robot instructions to further reveal learning processes unfolding during dynamic testing. This would further open ways to tailored assessment of individual children's potential for learning (Clabaugh, Ragusa, Sha, & Matarić, 2015; Granott, 2005) and more sophisticated understanding of children's differential development in ways that can directly impact upon their learning.

ACKNOWLEDGEMENTS

We would like to thank Bart Dirkx and Ruus van der Aalst for their help with programming their *robot-under-construction* for our study and Huguette Fles, Nathalie Ijsselmuiden, and Karlijn Nigg for their help in collecting the data for this study.

NOTE

¹This procedure was explained in the informed consent letter, ethically approved by our university, and a data protection compliance protocol was followed. All data were anonymized, and video-materials destroyed after coding and controlling the data.

ORCID

Wilma C.M. Resing  <https://orcid.org/0000-0003-3864-4517>

REFERENCES

Alexander, P. A., White, C. S., Haensly, P. A., & Crimmins-Jeanes, M. (1987). Training in analogical reasoning. *American Educational Research Journal*, 24, 387–404. <https://doi.org/10.3102/00028312024003387>

Alexander, P. A., Willson, V. L., White, C. S., & Fuqua, J. D. (1987). Analogical reasoning in young children. *Journal of Educational Psychology*, 79, 401–408. <https://doi.org/10.1037/0022-0663.79.4.401>

André, V., Jost, C., Hausberger, M., Le Pévédić, B., Jubin, R., Duhaut, D., & Lemasson, A. (2014). Ethorobotics applied to human behaviour: Can animated objects influence children's behaviour in cognitive tasks? *Animal Behaviour*, 96, 69–77. <https://doi.org/10.1016/j.anbehav.2014.07.020>

Baxter, P., Ashurst, A., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS ONE*, 12(5), 1, e0178126–23. <https://doi.org/10.1371/journal.pone.0178126>

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., ... Looije, R. (2013). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2), 33–53. <https://doi.org/10.5898/jhri.1.2.belpaeme>

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, 3, eaat5954. <https://doi.org/10.1126/scirobotics.aat5954>

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

Bethge, H., Carlson, J. S., & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven matrices performance, visual search behaviour, test anxiety and test orientation. *Intelligence*, 6, 89–97. [https://doi.org/10.1016/0160-2896\(82\)90022-8](https://doi.org/10.1016/0160-2896(82)90022-8)

Brown, L., & Howard, A. M. (2013, October). Engaging children in math education using a socially interactive humanoid robot. In *Humanoid Robots (Humanoids), 2013 13th IEEE-RAS International Conference* (pp. 183–188). IEEE.

Campione, J. C., & Brown, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment: An interactional approach to evaluating learning potential* (pp. 82–109). New York: Guilford Press.

Chang, C. W., Lee, J. H., Chao, P. Y., Wang, C. Y., & Chen, G. D. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Educational Technology & Society*, 13, 13–24.

Chin, K. Y., Hong, Z. W., & Chen, Y. L. (2014). Impact of using an educational robot-based learning system on students' motivation in elementary education. *Learning Technologies, IEEE Transactions on*, 7, 333–345. <https://doi.org/10.1109/TLT.2014.2346756>

Clabaugh, C., Ragusa, G., Sha, F., & Matarić, M. (2015). Designing a socially assistive robot for personalized number concepts learning in preschool children. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (pp. 314–319). IEEE.

Collins, M., & Laski, E. V. (2015). Preschoolers' strategies for solving visual repeating pattern tasks. *Early Childhood Research Quarterly*, 31, 204–214.

Csapó, B. (1997). The development of inductive reasoning: Cross-sectional assessments in an educational context. *International Journal of Behavioural Development*, 20, 609–626. <https://doi.org/10.1080/016502597385081>

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies – why and how. *Knowledge-Based Systems*, 6, 258–266. [https://doi.org/10.1016/0950-7051\(93\)90017-N](https://doi.org/10.1016/0950-7051(93)90017-N)

Deublein, A., Pfeifer, A., Merbach, K., Bruckner, K., Mengelkamp, C., & Lugrin, B. (2018). Scaffolding of motivation in learning using a social robot. *Computers & Education*, 125, 182–190. <https://doi.org/10.1016/j.comedu.2018.06.015>

Elliott, J. G., Grigorenko, E. L., & Resing, W. C. M. (2010). Dynamic assessment: The need for a dynamic approach. In P. Peterson, E. Baker, & B. McGaw (Eds.), *International Encyclopedia of Education* (Vol. 3) (pp. 220–225). Amsterdam, The Netherlands: Elsevier.

Freund, P. A., & Holling, H. (2011). How to get really smart: Modeling retest and training effects in ability testing using computer-generated figural matrix items. *Intelligence*, 39, 233–243. <https://doi.org/10.1016/j.intell.2011.02.009>

Granott, N. (2005). Scaffolding dynamically toward change: Previous and new perspectives. *New Ideas in Psychology*, 23, 140–151. <https://doi.org/10.1016/j.newideapsych.2006.07.002>

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities*, 42, 111–132. <https://doi.org/10.1177/0022219408326207>

Guthke, J., & Beckmann, J. F. (2000). The learning test concept and its application in practice. In C. S. Lidz, & J. G. Elliott (Eds.), *Advances in*

cognition and educational practice (Vol. 6). *Dynamic Assessment: Prevailing models and applications*. (pp. 17–69). New York: Elsevier.

Hessels-Schlatter, C. (2002). A dynamic test to assess learning capacity in people with severe impairments. *American Journal on Mental Retardation*, 107, 340–351. [https://doi.org/10.1352/0895-8017\(2002\)107<0340:ADTTAL>2.0.CO;2](https://doi.org/10.1352/0895-8017(2002)107<0340:ADTTAL>2.0.CO;2)

Holzman, T. G., Pellegrino, J. W., & Glaser, R. (1983). Cognitive variables in series completion. *Journal of Educational Psychology*, 75, 603–618. <https://doi.org/10.1037/0022-0663.75.4.603>

Hong, Z. W., Huang, Y. M., Hsu, M., & Shen, W. W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL Classrooms. *Journal of Educational Technology & Society*, 19, 337–349.

Huang, S. H., Wu, T. T., Chu, H. C., & Hwang, G. J. (2008, March). A decision tree approach to conducting dynamic assessment in a context-aware ubiquitous learning environment. In *Proc. Fifth IEEE International Conference on Wireless, Mobile, and Ubiquitous Technology in Education*, 89–94.

Hussain, S. L., Lindh, J., & Shukur, J. G. (2006). The effect of LEGO training on pupils' school performance in mathematics, problem solving ability and attitude: Swedish data. 2006. *Educational Technology & Society*, 9, 182–194.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2011). Making instruction and assessment responsive to diverse students' progress: Group-administered dynamic assessment in teaching mathematics. *Journal of Learning Disabilities*, 44, 381–395. <https://doi.org/10.1177/0022219411407868>

Jones, A., & Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics*, 10, 357–370. <https://doi.org/10.1007/s12369-017-0458-z>

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Aylin, C., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives*, 12(3), 146–151. <https://doi.org/10.1111/cdep.12277>

Khandelwal, M. (2006). Teaching table: A tangible mentor for pre-kindergarten math education. Unpublished master's thesis, Georgia Tech University, Atlanta, Georgia.

Khandelwal, M., & Mazalek, A. (2007). Teaching table: A tangible mentor for pre-k math education. In *1st International Conference on Tangible and Embedded Interaction (TEI'07)* (pp. 191–194). <https://doi.org/10.1145/1226969.1227009>

Klauer, K. J., & Phye, G. D. (2008). Inductive reasoning: A training approach. *Review of Educational Research*, 78, 85–123. <https://doi.org/10.3102/0034654307313402>

Kossowska, M., & Nęcka, E. (1994). Do it your own way: Cognitive strategies, intelligence, and personality. *Personality and Individual Differences*, 16, 33–46. [https://doi.org/10.1016/0191-8869\(94\)90108-2](https://doi.org/10.1016/0191-8869(94)90108-2)

Kozima, H., & Nakagawa, C. (2007). A robot in a playroom with preschool children: longitudinal field practice. Proceedings of 16th IEEE International Conference Robot & Human Interactive Communication (ROMAN 2007), 16, 1058–1059.

Liu, M., Scordino, R., Geurtz, R., Navarrete, C., Ko, Y., & Lim, M. (2014). A look at research on mobile learning in K-12 education from 2007 to the present. *Journal of Research on Technology in Education*, 46, 325–372. <https://doi.org/10.1080/15391523.2014.925681>

Manches, A., & O'Malley, C. (2016). The effect of physical manipulatives on children's numerical strategies. *Cognition and Instruction*, 34, 27–50. <https://doi.org/10.1080/07370008.2015.1124882>

Molnár, G., Greiff, S., & Csapó, B. (2013). Inductive reasoning, domain specific and complex problem solving: Relations and development. *Thinking Skills and Creativity*, 9, 35–45. <https://doi.org/10.1016/j.tsc.2013.03.002>

Moriguchi, Y., Kanda, T., Ishiguro, H., Shimada, Y., & Itakura, S. (2011). Can young children learn words from a robot? *Interaction Studies*, 12, 107–118. <https://doi.org/10.1075/is.12.1.04mor>

Movellan, J. R., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI 2009)* (pp. 307–308). <https://doi.org/10.1145/1514095.1514189>. La Jolla, CA, USA: ACM/IEEE.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1, 209–0015.

Passig, D., Tzuriel, D., & Eshel-Kedmi, G. (2016). Improving children's cognitive modifiability by dynamic assessment in 3D Immersive Virtual Reality environments. *Computers & Education*, 2016(95), 296–308.

Perret, P. (2015). Children's inductive reasoning: Developmental and educational perspectives. *Journal of Cognitive Education and Psychology*, 14, 389–408. <https://doi.org/10.1891/1945-8959.14.3.389>

Piaget, J. (1955). *The child's construction of reality*. London: Routledge & Kegan Paul Limited.

Raven, J., Raven, J. C., & Court, J. H. (2000). *Standard progressive matrices*. Oxford, England: Psychology Press.

Raven, J., Raven, J. C., & Court, J. H. (2003). *Manual for Raven's progressive matrices and vocabulary scales. Section 1: General overview*. San Antonio, TX: Harcourt Assessment.

Renninger, K. A., & Granott, N. (2005). The process of scaffolding in learning and development. *New Ideas in Psychology*, 23, 111–114. <https://doi.org/10.1016/j.newideapsych.2006.07.003>

Resing, W. C. M. (1993). Measuring inductive reasoning skills: The construction of a learning potential test. In J. H. M. Hamers, K. Sijtsma, & A. J. J. M. Ruijssenaars (Eds.), *Learning potential assessment. Theoretical, methodological and practical issues* (pp. 219–241). Amsterdam/Berwyn P.A., The Netherlands: Swets & Zeitlinger Inc.

Resing, W. C. M. (2000). Assessing the learning potential for inductive reasoning in young children. In C. S. Lidz, & J. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (pp. 229–262). Oxford, UK: Elsevier Inc.

Resing, W. C. M., Bakker, M., Pronk, C. M. E., & Elliott, J. G. (2017). Progression paths in children's problem solving: The influence of dynamic testing, initial variability, and working memory. *Journal of Experimental Child Psychology*, 153, 83–109. <https://doi.org/10.1016/j.jecp.2016.09.004>

Resing, W. C. M., & Elliott, J. G. (2011). Dynamic testing with tangible electronics: Measuring children's change in strategy use with a series completion task. *British Journal of Educational Psychology*, 81, 579–605. <https://doi.org/10.1348/2044-8279.002006>

Resing, W. C. M., Steijn, W. M. P., Xenidou-Dervou, I., Stevenson, C. E., & Elliott, J. G. (2011). Computerized dynamic testing: A study of the potential of an approach using sensor technology. *Journal of Cognitive Education and Psychology*, 10(2), 178–194. <https://doi.org/10.1891/1945-8959.10.2.178>

Resing, W. C. M., Touw, K. W. J., Veerbeek, J., & Elliott, J. G. (2017). Progress in the inductive strategy-use of children from different ethnic backgrounds: A study employing dynamic testing. *Educational Psychology*, 37(2), 173–191. <https://doi.org/10.1080/01443410.2016.1164300>

Resing, W. C. M., Xenidou-Dervou, I., Steijn, W. M. P., & Elliott, J. G. (2012). A "picture" of children's potential for learning: Looking into strategy changes and working memory by dynamic testing. *Learning and Individual Differences*, 22, 144–150. <https://doi.org/10.1016/j.lindif.2011.11.002>

Serholt, S., & Barendregt, W. (2016, October). Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (p. 64). ACM.

Serholt, S., Basedow, C. A., Barendregt, W., & Obaid, M. (2014, November). Comparing a humanoid tutor to a human tutor delivering an instructional task to children. In *Humanoid Robots (Humanoids), 2014 14th IEEE-RAS International Conference on* (pp. 1134–1141). IEEE.

Shinozawa, K., Naya, F., Yamato, J., & Kogure, K. (2005). Differences in effect of robot and screen agent recommendations on human decision-making. *International Journal of Human Computer Studies*, 62, 267–279. <https://doi.org/10.1016/j.ijhcs.2004.11.003>

Siegler, R. S., & Svetina, M. (2006). What leads children to adopt new strategies? A microgenetic/cross-sectional study of class inclusion. *Child Development*, 77, 997–1015. <https://doi.org/10.1111/j.1467-8624.2006.00915.x>

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. New York: Cambridge University Press.

Stevenson, C. E., Heiser, W. J., & Resing, W. C. M. (2013). Working memory as a moderator of training and transfer of analogical reasoning. *Contemporary Educational Psychology*, 38, 159–169. <https://doi.org/10.1016/j.cedpsych.2013.02.001>

Stevenson, C. E., Touw, K. W. J., & Resing, W. C. M. (2011). Computer or paper analogy puzzles: Does assessment mode influence young children's strategy progression? *Educational and Child Psychology*, 28, 2, 67–84.

Sullivan, F. R. (2008). Robotics and science literacy: Thinking skills, science process skills and systems understanding. *Journal of Research in Science Teaching*, 45, 373–394. <https://doi.org/10.1002/tea.20238>

Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the USA*, 104, 17954–17968.

Timms, M. J. (2016). Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms. *International Journal of Artificial Intelligence in Education*, 26, 701–712. <https://doi.org/10.1007/s40593-016-0095-y>

Tzuriel, D. (2001). *Dynamic assessment of young children* (pp. 63–75). New York: Kluwer/Plenum Press. <https://doi.org/10.1007/978-1-4615-1255-4>

Tzuriel, D., & Flor-Maduel, H. (2010). Prediction of early literacy by analogical thinking modifiability among kindergarten children. *Journal of Cognitive Education and Psychology*, 9, 207–226. <https://doi.org/10.1891/1945-8959.9.3.207>

Tzuriel, D., & George, T. (2009). Improvement of analogical reasoning and academic achievement by the Analogical Reasoning Programme (ARP). *Educational and Child Psychology*, 26, 71–93.

Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessment. *Journal of Computer Assisted Learning*, 18, 21–32. <https://doi.org/10.1046/j.0266-4909.2001.00204.x>

Veerbeek, J., Vogelaar, B., Verhaegh, J., & Resing, W. C. M. (2019). Process assessment in dynamic testing using electronic tangibles. *Journal of Computer Assisted Learning*, 35, 127–142.1-16. <https://doi.org/10.1111/jcal.12318>

Verhaegh, J., Fontijn, W., Aarts, E., & Resing, W. C. M. (2013). In-game assessment and training of nonverbal cognitive skills using TagTiles. *Personal and Ubiquitous Computing*, 17, 1637–1646. <https://doi.org/10.1007/s00779-012-0527-0>

Vogelaar, B., & Resing, W. C. M. (2016). Gifted and average-ability children's progression in analogical reasoning in a dynamic test setting. *Journal of Cognitive Education and Psychology*, 15, 349–367. <https://doi.org/10.1891/1945-8959.15.3.349>

Vogelaar, B., & Resing, W. C. M. (2018). Changes over time and transfer of analogy-problem solving of gifted and non-gifted children in a dynamic testing setting. *Educational Psychology*, 38(7), 898–914. <https://doi.org/10.1080/01443410.2017.1409886>

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

WittyWorX. (2012). WittyWorX. <http://www.wittyworx.com/index>

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17, 89–100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>

Wood, D., & Wood, H. (1996). Vygotsky, tutoring and learning. *Oxford Review of Education*, 22, 5–16. <https://doi.org/10.1080/0305498960220101>

Wu, H. M., Kuo, B. C., & Wang, S. C. (2017). Computerized Dynamic Adaptive Tests with Immediately Individualized Feedback for Primary School Mathematics Learning. *Educational Technology & Society*, 20, 61–72.

How to cite this article: Resing WCM, Bakker M, Elliott JG, Vogelaar B. Dynamic testing: Can a robot as tutor be of help in assessing children's potential for learning? *J Comput Assist Learn*. 2019;35:540–554. <https://doi.org/10.1111/jcal.12358>

APPENDIX A

SCHEMATIC OVERVIEW GENERAL INSTRUCTION PROVIDED: EXAMINER ACTIONS, CHILD ACTIONS AND OUTPUT SOURCES

TL-TT INTERACTION	INPUT	OUTPUT: VOICE ROBOT+C1	OUTPUT: VOICE FILE	OUTPUT: SOUND FILE	OUTPUT: ROBOT BEHAVIOUR
TL: instruction	TL: stamp				Eyes: Happy Body: Dancing
TL: shows the basket	TL: stamp	Hello, I am MyRo. Today, we are going to make some puzzles, I will tell you what we have to do.		SndDisco1.wav	
		Here we have baskets with puzzle pieces. You can make a puppet with them, here on this plate.	VoiceIntroductionA.wav		Eyes: Happy
		The puppet is made of different pieces: A head, a tommy (made of 3 pieces). Arms. And legs.	VoiceIntroductionB.wav		Eyes: Neutral
		A lot of different puppets can be made. It can be a boy or a girl.	VoiceIntroductionC.wav		Eyes: Neutral
		There are four colours: Pink, Yellow, Green. And blue.	VoiceIntroductionD.wav		Eyes: Neutral
		The pieces can have strips, dots, or nothing.	VoiceIntroductionE.wav		Eyes: Neutral
			VoiceIntroductionF.wav	SndBoardEmpty.wav	Eyes: Neutral
	IF: piece on board	First, empty the plate and put the pieces in the baskets.	VoiceBoardEmpty.wav		Eyes: Neutral
		And now, make a nice puppet on the plate. If you are ready you can softly push my head.	VoiceIntroductionG.wav		Eyes: Happy Body: Neutral
TL: helps	TT: solve puppet		VoiceTTStamp.wav		
	IF: piece outside puppet	Place the pieces within the lines	VoicePuppetField.wav	SndPuppetField.wav	
	TT: stamp				
	IF: puppet not complete	The puppet is not yet finished	VoicePuppetFinish.wav	SndPuppetFinish.wav	Eyes: Neutral Body: Shaking head (no)
	IF: puppet complete	Well done!	VoiceGood.wav	SndGood.wav	Eyes: Happy Body: Nodding (yes)
			VoiceGood.wav		

TL	Action examiner	
	If something happens	Task not solved
	Action child	Task solved