

The Young Child's Cognition

Citation for published version (APA):

Urlings, C. C. J. (2020). The Young Child's Cognition: Intervention and innovative measurement. [Doctoral Thesis, Maastricht University]. Maastricht University. <https://doi.org/10.26481/dis.20201030cu>

Document status and date:

Published: 01/01/2020

DOI:

[10.26481/dis.20201030cu](https://doi.org/10.26481/dis.20201030cu)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

THE YOUNG CHILD'S COGNITION

Intervention and innovative measurement

Corrie Urlings

© copyright Corrie C.J. Urlings, Maastricht 2020

Printing: ProefschriftMaken || www.proefschriftmaken.nl

Cover Design: Pixelkings & Corrie Urlings

ISBN 978 94 6380 774 6

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of the author or the copyright-owning journals for previous published chapters.

THE YOUNG CHILD'S COGNITION

Intervention and innovative measurement

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,

op gezag van de Rector Magnificus, Prof. dr. Rianne M. Letschert

volgens het besluit van het College van Decanen,

in het openbaar te verdedigen

op vrijdag 30 oktober 2020 om 12.00 uur

door

Corrie Christina Jeannine Urlings

Promotor

Prof. dr. L. Borghans

Copromotor

Dr. K.M. Coppens

Beoordelingscommissie

Prof. dr. B.H.H. Golsteyn (voorzitter)

Prof. dr. K. Moeller (Loughborough University, United Kingdom)

Prof. dr. W.C.M. Resing (Leiden University)

Prof. dr. T. Schils

Acknowledgements

Over the last 4.5 years, I gained many things for which I am very thankful. A renovated home, an art school degree, a husband, two cats, experience with traveling abroad by myself, a lot of 'wisdom' (which came with a lot of wrinkles) and a finished PhD.

Perhaps I should start at what I consider the beginning of my love for research, when my curiosity was first triggered. In September 2006 I started my bachelor studies in psychology at Maastricht University. Those years in Maastricht I have experienced as a time of intellectual growth. The doors to the complexity of human behavior were opened to me and I loved every second of it. I therefore returned to Maastricht University for a junior researcher position with great excitement in 2015. In that same year I started a PhD position in which I could do an innovative, multidisciplinary project on my passion, which is the measurement of cognitive skills in children. Whilst my time as an undergraduate student was a time of intellectual growth, my PhD years I consider to be a time of personal growth. Managing my own complicated project, setting up diverse collaborations and learning how to teach, whilst having multiple hobbies and projects outside of work was more challenging and rewarding than I could ever imagine. For all these reasons, I feel a close connection to Maastricht, its university and the people I have met during these times.

Having worked outside of academia before I started my career in research, I believe it is true what many people say, that academia is like a bubble. Although this phrase may have a negative connotation to some, I feel more inclined toward a positive interpretation. It is a bubble filled with people who are curious and proactive, who dare ask questions and work (very) hard to find answers. Even though we are confronted with different opinions every day, there is a strong sense of unity in the scientific community. This I have not only experienced on campus, but also while visiting other universities or when meeting people at (inter)national conferences. Needless to say, the inspiration that can be found is endless and the atmosphere is quite unique.

But to only thank academia or the university as a whole would not do justice to the amount of support, patience and guidance I received from several people in particular during my PhD. My supervisors, Lex and Karien, have continuously supported me through the *maze* these projects were. I received so many opportunities from them to shape the projects and my own

personal development in any way I could manage. To Lex, thank you for always providing new perspectives on my ideas, findings and problems. You were always encouraging me to think outside the box in the projects, which is exactly what was needed to complete it successfully. To Karien, you were always open to a conversation to discuss any topic. Dealing with the complexity and the stress it can bring at times, you had big and small pieces of advice for everything. Thank you for your help, encouragement and giving me the feeling you had complete faith in me.

I have had the privilege to collaborate with people from different disciplines on the projects. These people deserve special mentioning. To Rico, thank you for collaborating on this project and believing in its value. I enjoyed the collaboration with the Swarm Lab, not only because of the people, but also because a visit entailed admiring all those fancy robots and drones. To Seethu, who worked relentlessly with me on the data: our meetings and many e-mails have resulted in a great research projects. You as well as Lukas were also great to work with during the Hackathon in Heidelberg, where we raced against time to generate a solution on very little sleep in hot weather. To Petra, thank you as well. Your continuous helpful suggestions helped shape my thoughts and ideas. I presented my research ideas and my questions one time during a seminar at the Oxfordlaan and you were suggested by the audience as a source of help. I am very glad I followed up on that suggestion. I also want to thank the people of IDEE who have collaborated on the project by building the amazing tool we used for the studies. Special mention goes to Paul, Henri and René for their efforts in the project. They say advice is always given to the ones who ask for it, so thank you, Trudie, for the insights you gave me with great personal flair in the last months of my PhD regarding time and project management. And finally, a very important partner: the schools that allowed me to collect data. All teachers and principals as well as André and Fons from Kindante and STEM-II, who facilitated all communication, a huge thank you for your trust and collaboration.

I have met so many wonderful people in the last few years with whom I have shared an office, a lunch, interesting talks or walks. In particular from 'TS6', a big thank you to Annemarie, Anne, Elise, Jacqueline, Liudmila, Mariana, Marjolein, Max, Mélanie, Menw, Olga, Raoul, Ron, Roxanne, Sergio, Stan, Stefa, Tyas and Vanessa, who have each contributed in their own specific way to my time as a PhD student. A special mention should of course be made of Britt and Lin. Britt, we met thanks to Tyas right when I started working at the university and we remained friends all this time. Thank you for the many lunches,

walks and coffee breaks we shared over the years and for always putting my doubts into perspective. Lin, your enthusiasm and encouragement of 'women in science' was and continues to be an inspiration to me. Whenever I am unsure about my direction, I know I can always count on your encouragement to take the leap. Thanks to Fleur, Marion, Silvana and Sylvia, who were always ready to answer questions and provide help when needed. I also want to mention the Pomodoro writing group with great fondness, specifically Inge, Kim, Maja, Raginee and Sanne. I had the pleasure of joining the writing days in the final year of my PhD and I am very glad I did, since I always experienced renewed energy thanks to this social circle. Regarding that last phase, I want to say thank you to the wonderful colleagues at the department of Educational Research and Development as well for celebrating the successful completion of my PhD with me.

Last but not least, my family, family-in-law, and friends. Aan mijn mama, je bent het ultieme voorbeeld van wilskracht en doorzettingsvermogen, merci voor je voortdurende aanmoediging. To my brother Peter, we are close no matter how far! When I think of you, I think of curiosity and science. You have influenced me more than you may think. To the lovely ladies Eline, Sofie and Ariane, thanks for the lovely shopping trips, dinners, pizza and/or movie nights that kept me sane. And then there is my dear husband Chris, who celebrated every success with me and put every problem in perspective, who always believed in my skills and the successful completion of the projects. We even had the opportunity to do a project together in Germany, which was great. Your support was invaluable to me. I feel they should hand out certificates of bravery to the spouses of PhD students, because you definitely deserve one.

Corrie Urlings, March 2020.

Table of contents

Chapter 1: Introduction	1
1.1 Motivation	2
1.2 Aim	4
1.3 Outline	5
1.4 References	9
Chapter 2: Does preschool education influence cognitive outcomes? A closer look at the Netherlands	11
2.1 Introduction	12
2.1.1 Brian development and early life experiences	13
2.1.2 Effectiveness of preschool education international	14
2.1.3 Early childhood education in the Netherlands	15
2.2 Method	19
2.2.1 Data	19
2.2.2 Sampling procedure	21
2.2.3 Data analyses	21
2.3 Results	22
2.3.1 Descriptives of the full sample	22
2.3.2 Descriptives of target group sample	24
2.3.3 Multiple regression analysis	25
2.3.4 Conclusion of results	28
2.4 Discussion	30
2.5 Conclusion	33
2.6 References	34
Chapter 3: Measurement of executive functioning using a playful robot in kindergarten	41
3.1 Introduction	42
3.1.1 History and the current use of robots in education	42

3.1.2 Understanding executive functioning in kindergarteners	42
3.1.3 Using robots to stimulate executive functions	43
3.1.4 Using robots to measure executive functions	44
3.1.5 The Bee-Bot	44
3.1.6 The current study	45
3.2 Method	46
3.2.1 Participants	46
3.2.2 Tasks, instruments and measurements	47
3.2.3 Procedure	50
3.2.4 Analyses	51
3.3 Results	51
3.3.1 Descriptive statistics	51
3.3.2 Confirmatory factor analysis	53
3.4 Discussion	56
3.4.1 Interpretation of results	56
3.4.2 Implications of the findings	57
3.4.3 Limitations of the study	58
3.5 Conclusion	59
3.6 References	60
Chapter 4: A digital wooden tabletop maze for estimation of cognitive capabilities in children	65
4.1 Introduction	66
4.2 Digital wooden maze: Hardware and data processing	68
4.3 Digitalization of pen positions through Computer Vision	72
4.4 Experiments and results	73
4.5 Discussion and future work	76
4.6 References	77
Chapter 5: Children’s executive functioning: What can solving mazes using Clever Maze reveal?	81
5.1 Introduction	82

5.1.1 The differentiation of executive functions in children	82
5.1.2 Executive functioning tests and the role of technology	83
5.1.3 The current study: Sensors in a maze toy	85
5.2 Method	86
5.2.1 Participants	86
5.2.2 Procedure	86
5.2.3 Analyses	94
5.2.4 Ethics	94
5.3 Results	94
5.3.1 Descriptive statistics	94
5.3.2 Correlational analyses	96
5.3.3 Factor analyses	98
5.4 Discussion	99
5.4.1 Interpretation of the results	100
5.4.2 Implications	102
5.4.3 Limitations and recommendations	103
5.5 References	105
Chapter 6: Free play and executive functioning: What can Clever Maze reveal?	111
6.1 Introduction	112
6.1.1 The role and assessment of play	113
6.1.2 The role of technology	114
6.1.3 Cognitive skills and assessments in kindergarten	114
6.1.4 The current study	115
6.2 Method	116
6.2.1 Participants	116
6.2.2 Procedures and materials	116
6.2.3 Data and analyses	119
6.3 Results	119
6.3.1 Descriptives	119

6.3.2 Correlational analyses	120
6.4 Discussion	123
6.4.1 Limitations and recommendations	125
6.4.2 Implications of findings	125
6.5 References	127
Chapter 7: Conclusion and valorisation	131
7.1 Conclusion	132
7.2 Implications and valorisation	135
7.2.1 The relation between preschool education and cognitive outcomes on standardized school tests	135
7.2.2 The relation between executive functioning skills and programming a Bee-Bot	135
7.2.3 Development of a smart toy as a measurement of executive functioning skills	136
7.2.4 Children's executive functioning: What can solving mazes and free play on Clever Maze reveal?	136
7.3 References	139
Appendices	141
Appendix A	141
Appendix B	145
Appendix C	150
Nederlandse samenvatting (Dutch summary)	155
About the author	163



Chapter 1

Introduction

1.1 Motivation

Research findings have highlighted the importance of early childhood cognitive skills and development. Furthermore, findings on the flexibility of the brain have adjusted the long-standing image of the brain being fully developed at the time of birth. Neurological structural and functional malleability of the young brain (Brown & Jernigan, 2012) implies a prominent role of the environment for shaping the brain in early childhood. For instance, the long-lasting effects of negative early life experiences have been well documented in the case of adverse situations, such as low socioeconomic status (Bradley & Corwyn, 2002; Brito & Noble, 2014). With the increased realization of responsibility for providing stimulating childhoods, more emphasis has been placed on understanding and stimulating cognitive skills, even before children enter formal primary education. However, in order to fully understand whether interventions are effective or not, that is, whether children have improved their skills or not, proper assessment tools are very important. Not only should early testing allow an understanding of the current developmental level of a child, comparisons among children should be possible, as well as tracking over time, allowing insight into skill growth or evolution. These tests should provide insights into skills in a valid and reliable way, which often requires standardized conditions and a tendency toward strenuous testing conditions for which much expertise and time is needed from the examiner. The younger the child, however, the more difficult these conditions can be realized.

A multitude of tests are available for measuring various cognitive skills. For different developmental stages, there are varying ranges of age-appropriate tools and tests for specific skills available, which is imperative given the vast variety of skills between age groups. Each test has its own advantages and disadvantages, which often lead to continuous development of tests. For older children, options are ample once they have attained sufficient self-regulation and vocabulary skills to successfully endure rigorous testing. However, the younger children are, the harder it is to develop a proper test that is valid and reliable. A particularly interesting group are kindergarteners, as they are still developing the cognitive skills needed for steering everyday behaviour, yet gaining insight into their skills is important as they are soon to start formal primary education. For them, strenuous, long-lasting testing situations are (too) demanding. Rather than formal tests, other forms of testing can be used such as (systematic) observation (scales) or surveys and questionnaires. Though they provide more insight into

everyday behaviour of the children, these forms of testing require large time investments to perform as well as expertise for administration and scoring, similar to formal testing situations.

It seems the activity of standardized testing itself is not in line with the needs and daily lives of kindergarteners. The predominant activity of kindergarteners is playing (Hofferth & Sandberg, 2001), which seems to be in contrast to the requirements of formal testing situations. However, the importance of play is not trivial. Studies have shown that play holds a crucial role in child development (Bodrova, 2008; Wood, 2014). Even curricula in kindergarten are often taught via play or games. For many professionals, the way the testing culture has taken shape causes friction with what is generally perceived as the main activity of young, learning children, namely play (Werkgroep en Steungroep Kleuteronderwijs, 2013). Therefore, it is warranted to investigate whether a more playful approach can be used when testing young children.

The focus of this thesis is on kindergarteners, the importance of their developmental stage and how cognitive abilities could be measured keeping the earlier described restrictions as well as possibilities in mind for this age group. The aim is to work toward an assessment approach that focuses more on playful behaviour of children and how this behaviour can be measured, using innovative tools. Given the value of play, this thesis will examine the possibilities of measuring play behaviour further by using technological means. If we can create tools that are smart enough to collect data on their own, the burdens of existing tools could potentially be lessened. In the case of kindergarteners, it makes most sense to use toys as our tool of choice for tackling this question. While children play, they do this according to their level of cognitive development, which can be observed and quantified, as is often done when using observations and/or questionnaires. Obtaining a comprehensive view on child cognitive ability is important. Depending on the purpose of the testing, observation can be used in combination with other standardized tests. This approach, if successful, would imply testing can be done without testing, i.e. by letting children play whilst the toys collect data on cognitive abilities. A computer or tablet task is to some extent a good solution as it can collect data on its own, however, screen-based task solving may affect reliability, ecological validity, applicability and other biases in comparison to pencil-and-paper tasks (Csapó, Molnár, & Nagy, 2014).

In a world where data are continuously collected and more attention is given to ‘big data’ and ‘the internet of things’, surely technological applications can pose alternative and feasible solutions for testing of groups that are difficult to test and to collect data on. Technological innovations are currently often designed and incorporated as a means for cognitive stimulation, such as interactive robots or tablet games designed to improve specific skills. These means could however also be used for detailed data collection and increasing insights into the current skill level. This is for example done in stealth assessment. In this type of assessment, learner performance data are continuously gathered during playing and/or learning. Next, inferences are made about the competency level of the learner and this information is then used to support the learning (Shute, Ventura, Bauer & Zapata-Rivera, 2009). As technological applications can be more embedded in real-life than ever before, natural behaviour can be measured rather than the more artificial setting of pencil-and-paper or screen-based task making. Therefore, in this thesis, the use of a programmable robot and the integration of sensor technology into a toy is chosen as a means of data collection. This approach ensures that the children interact with a tool (i.e. toy) that exists in the space as well as collect valuable data on their play behaviour.

1.2 Aim

There is a need for innovative developments to solve friction between standard testing situations and cognitive ability and play in kindergarteners. First, this thesis will underline the importance of attention provided to young children by studying the effectiveness of a cognitive intervention in kindergarteners. Therefore, the effects of a preschool education program in the Netherlands in relation to standardized, academic achievement tests in kindergarten will be examined first. The remainder of the thesis will focus on issues of kindergartener assessment using play within the framework of executive functions (EF). Central to the approach of this thesis is to include technological means in comparison to traditional EF tests that are not screen-based tasks on computers or tablets in order to use real-life materials, while still maintaining a reduction in time and expertise investments. For this study, the Bee-Bot was utilized. This small programmable robot is available for purchase and many primary schools (teachers) are familiar with it or with the concept. The contribution of this study is knowledge about how or to what extent children’s manipulation of this tool whilst solving tasks can provide insight into EF. Next, we utilized a new tool that looks like a regular toy and explored its use as an assessment tool, i.e. to provide insight into EF. The toy is smart as sensor technology is

embedded into it, allowing data collection on the behaviour of the children. This assessment tool, called Clever Maze (CM), was tested in both a structured setting similar to traditional EF tests and an unstructured setting similar to play. Data collected with CM were compared with children's outcomes on existing, traditional EF tests to understand what information can be collected with CM. The comparison of the outcomes on these measurements yields valuable insights into the type of information that is collected with the tool. Specific attention is provided to information collected on processes of solving maze assignments. Specifically, the work where children play with CM thus provides insights into the question of whether “testing without testing” is at all possible. The studies on the smart toy serve to increase the understanding technological advances can play in assessment of EF in a population that is difficult to test, namely young children.

Based on the different chapters, we answer the main research question of this thesis: *Given the importance of early measurement and tracking, as well as friction between everyday behaviour of kindergarteners and standardized tests, is it possible to devise assessments for kindergarteners that are more reliant on play and natural behaviour, utilizing technological applications?* Each chapter serves the following sub-questions respectively:

- *What is the relation between early childhood education and outcomes on standardized, academic achievement tests in Dutch kindergartens? (Chapter 2)*
- *What is the relation between executive functions and solving a task using a programmable robot in kindergarteners? (Chapter 3)*
- *How can we successfully build a sensor embedded toy for executive functions measurement in kindergarteners? (Chapter 4)*
- *What relations can be found between performance on traditional executive functions tests and making assignments on the Clever Maze? (Chapter 5)*
- *Can meaningful information on unstructured play, as collected by Clever Maze, yield insights on executive functions? (Chapter 6)*

1.3 Outline

The following provides an overview of the chapters of this thesis.

1.3.1 Chapter 2: The Relation Between Preschool Education and Cognitive Outcomes on Standardized School Tests

Given the importance of early childhood experiences, early intervention has great potential in combatting developmental gaps. However, researching the effectiveness of early education programs in the Netherlands has been difficult due to the many policy changes over the years, the decentralized system, and the dual system of daycare and playgroups. To overcome this problem, data from a large cohort study were used, which allowed controls for changes over time and in different regions, as well as for use of facilities and a preschool education program, and background factors such as socioeconomic status, language spoken at home, and parental country of birth. The results of multiple regression analyses indicated an overall positive relationship between having participated in preschool education programs and better educational outcomes for all children involved. When analyzing the target group children of preschool education programs, the relationship between following such a programme and cognitive outcomes was stronger as indicated by a higher coefficient. These coefficients were however less significant, probably due to smaller sample size. The unique data and our findings contribute to the national and international discussion regarding the role of preschool education in early childhood development.

1.3.2 Chapter 3: The Relation Between Executive Functioning Skills and Programming a Bee-Bot

We explored the potential of a robotics application in education as a measurement tool of child executive functioning skills. 65 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 (EF). The latent variable significantly predicted performance on a non-verbal ability task, a planning task, a verbal fluency task, and a memory task. It did not significantly predict pencil-and-paper tasks that measured visuospatial ability and non-verbal or design fluency. This study shows that it is possible to utilize a playful robotics task to obtain insightful information on child ability. We recommend more research on using diverse robots in larger samples with different age groups to explore the possibilities of robots as test instruments.

1.3.3 Chapter 4: Development of a Smart Toy as a Measurement of Executive Functioning Skills

Standardized tests play an important role in assessing a child's cognitive capabilities. The results of such tests are used e.g. in schools and kindergartens to analyze and support the development of the tested child. Classical standardized tests often result in high work load for the testing professionals who need to be present during the test, observe, document, and evaluate the test results. In addition, standardized tests typically rely on well-controlled environments that do not fully reflect real-life situations. Computer-based stealth-assessments that e.g. use a gaming environment to be fun and to hide the assessment from the child might present a valid alternative. However, for young children such as kindergartners, computer-based tests are not easily applicable due to technological boundaries. In this paper, we thus explore the alternative approach: physical game devices with a look and feel similar to toys typically provided to young children but that embed the electronics required for computer-based stealth testing. As a result, the game device – in our case a wooden tabletop maze – combines advantages of standardized computer-free and computer-based assessments. The device could allow for assessments in less structured environments without creating technological boundaries for the children.

1.3.4 Chapter 5: Children's Executive Functioning: What Can Solving Mazes Using Clever Maze Reveal?

Gaining insight into the EF of kindergartners is highly valuable. However, most tests measuring EF require standardized, strenuous testing conditions. In this study, we investigated the validity of an innovative approach to examine EF in children, i.e., a smart toy embedded with sensor technology, called Clever Maze (CM). CM looks on the exterior like a regular maze toy. In this study, 103 typically developing kindergartners (average 6 years old) made assignments with CM. Across these assignments, we created four factors out of all CM output extracted, using confirmatory factor analysis. These factors provide data on how children solve CM mazes, i.e., (1) the distance needed to solve mazes, (2) the total time standing still, (3) the average speed of moving and (4) the number of dead ends entered. The resulting factors were correlated to outcomes on traditional, paper-and-pencil EF tests, using Pearson's pairwise correlations to validate our main CM outcome measures. The results indicate significant relations between how children solve CM assignments and the traditional EF test, particularly

regarding visuospatial reasoning and visuospatial memory, problem-solving and to some extent working memory. The results are discussed in terms of the convergent and discriminant validity of the CM. The study illustrates how sensor technology could be used to collect specific information about children's EF and cognitive development. It also illustrates how technology can be employed in assessment with the use of real-life toys instead of computers or tablets.

1.3.5 Chapter 6: Free Play and Executive Functioning: What Can Clever Maze Reveal?

In early childhood, play is considered to be the driving factor of development. This view is widely accepted and most didactic materials embody this philosophy. However, when assessing young children's cognitive skills, standardized tests are used instead of play. In the study in the previous chapter, CM was used in a structured way where children made assignments with the toy and purposefully solved mazes. In this study, children can play freely with CM, whilst CM collects data on their behaviour. We extracted information on the total distance they travelled inside the maze, the time they played, the number of dead ends they entered as well as the time and distance travelled within those dead ends, their average speed of moving and the time they stood still. Spearman's non-parametric correlation was used to study the relation between these CM variables and outcomes on various, traditional EF tests, as well as parent's answers on a questionnaire inquiring on their child's EF. The results showed several significant correlations between CM and the EF tests, specifically with planning abilities, problem-solving and working memory. Between CM and the BRIEF there was a relation with taking initiative as well as flexibility. These findings are discussed in relation to convergent validity. There were no unexpected relations found regarding discriminant validity.

1.4 References

- Bodrova, E. (2008). Make-believe play versus academic skills: a Vygotskian approach to today's dilemma of early childhood education. *European Early Childhood Education Research Journal*, 16(3), 357-369. doi:10.1080/13502930802291777
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371-399.
- Brito, N. H., & Noble, K. G. (2014). Socioeconomic status and structural brain development. *Frontiers in neuroscience*, 8, 1-12.
- Brown, T. T., & Jernigan, T. L. (2012). Brain development during the preschool years. *Neuropsychology Review*, 22(4), 313-333. doi:10.1007/s11065-012-9214-1
- Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, 106(3), 639-650. doi:10.1037/a0035756
- Hofferth, S. L., & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family*, 63(2), 295-308.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge, Taylor and Francis.
- Werkgroep en Steungroep Kleuteronderwijs. (2013). *Zwartboek Kleuters in de knel!* Retrieved from <https://www.wsk-kleuteronderwijs.nl/wp-content/uploads/2013/04/Zwartboek-Kleuters-in-de-knel-april-2013-LR.pdf>
- Wood, E. A. (2014). Free choice and free play in early childhood education: Troubling the discourse. *International Journal of Early Years Education*, 22(1), 4-18.



Chapter 2

Does preschool education influence cognitive outcomes? A closer look at the Netherlands

This chapter is based on joint work with Karien Coppens^a and Lex Borghans^a.

^aDepartment of Macro, International and Labour Economics, School of Business
and Economics, Maastricht University

2.1 Introduction

A stimulating environment from an early age on, specifically targeted at at-risk groups of children and families, is of high importance for individual development, as well as society as a whole. Early childhood education programs are designed to target children who are at a higher risk of such a developmental disadvantage, which are usually linked to lower SES backgrounds, foreign languages spoken at home or an immigrant status (Burger, 2010). The aim of such programs is to narrow or even close the developmental gaps and provide children with equal opportunities before they enter formal primary education. Early childhood is an important period for brain development during which major functional and structural changes take place (Brown & Jernigan, 2012). Environmental influences can affect the early development of the brain via stimulation or deprivation (Hertzman & Boyce, 2010). Through gene-environment interactions, unique developmental trajectories unfold, allowing some children to thrive, while others have a risk of falling behind and these individual differences manifest early on in life. As Shonkoff and Phillips (2000) concluded in their work: “Striking disparities in what children know and can do are evident well before they enter kindergarten. These differences are strongly associated with social and economic circumstances, and they are predictive of subsequent academic performance” (p.5). There is growing awareness that education provided in schools is insufficient to bridge developmental gaps that have formed at such an early age (Driessen, 2004).

The effectiveness of Early Childhood Education (and Care) programs (ECEC) has been shown by targeted, intense experiments in the US such as Head Start and the Perry Preschool Project (Heckman, Moon, Pinto, Saveljev, & Yavitz, 2010; Schweinhart et al., 2005). A large meta-analysis of international research showed the effectiveness in other countries in quasi-experimental research settings (Burger, 2010). In the Netherlands, however, study findings up to now are mostly inconsistent (Fukkink, Jilink, & Oostdam, 2015, 2017). Researching the effectiveness of ECE programs in the Netherlands is difficult due to the many policy changes over the years, a decentralized system (allowing for differences in policy and practice between municipalities), and a dual system of daycare and preschool. In recent years, more attention is provided to ECE, with even a new law implemented in 2010 aiming to intensify policy and increase quality (Ministerie van Onderwijs, Cultuur en Wetenschap, 2015). The Inspectorate of Education reported on ECE numbers in the 37 big cities of the Netherlands in 2016, showing that there was a coverage of 71% ECE use by target group children in 2013 and 78% use in

2014 (Inspectie van het Onderwijs, 2016). A vast research population is needed to compare target-group children who followed an ECE program to those who did not in order to understand effectiveness.

To increase the understanding of the relationship between preschool education programs and cognitive outcomes in the Netherlands, we use a unique dataset from the Southern region of the Netherlands in which data were collected for several consecutive years in 18 municipalities. By studying the relationship between the provision of preschool education programs to children before their entry to kindergarten and their performance on an academic achievement test at the end of kindergarten, we aim to add to the knowledge on the effects of ECE programs in the Netherlands.

2.1.1 Brain Development and Early Life Experiences

To understand the potential impact of early childhood education, it is important to understand the mechanisms through which the outer environment can influence child development. Since the brain is still developing (Brown & Jernigan, 2012), particular variations in experiences children have influence the development differentially, leading to neurological changes (Noble, Tottenham, & Casey, 2005). Therefore, early childhood provides a unique window of opportunity because of a heightened capacity for change in the foundation of skill development and neural circuits (Knudsen, Heckman, Cameron, and Shonkoff, 2006). Due to heightened neuronal changeability, the brain is extra susceptible for external stimulation or deprivation during the early years, also known as 'sensitive periods' (Fox, Levitt, & Nelson, 2010; Knudsen, 2004). These periods exist on different ages for different levels of neural circuits. More specifically, the processing of stimuli is a lower-level circuit which ends around birth. But high-level circuits, such as higher cognitive functioning, have sensitive periods of several years after birth (Shonkoff, 2007). In a study by Kautz, Heckman, Diris, Ter Weel, and Borghans (2014), the authors analyzed the characteristics of effective early intervention programs and concluded that interventions before the age of 3 impacted cognitive abilities or IQ, whereas interventions that took place after this age mostly impacted non-cognitive skills. Since skill and brain development are hierarchical processes in which the higher levels depend and build on the lower levels, a good foundation is of great importance (Knudsen et al., 2006). Therefore, it is plausible to assume that the acquisition of (cognitive) skills at a young age during sensitive neurological periods begets the acquisition of more (cognitive) skills in later

development. As Knudsen et al. (2006) stated (p. 10161), “skills beget skills, success begets success”. Given that the environment plays a significant role in child development, there is high potential for mediating development via environmental interventions (Noble et al., 2005). An important example of an environmental factor is discussed by (Bradley & Corwyn, 2002), namely socioeconomic status (SES). Differences in SES lead to differences in services, goods, parental actions and social connections offered to the children. The effects of SES are therefore widespread, and among other outcomes also manifest in academic achievement outcomes. Researchers generally acknowledge that less-educated parents are lacking in the stimulation of young children’s development (UNESCO, 2015). Not necessarily unconnected to SES, other environmental factors may play an important role in preventing child developmental delays, such as family immigration and (quality of) language stimulation at home. Not only do children learn a second language best up to a certain age, but even native children need sufficient language stimulation in order to avoid developmental problems.

2.1.2 Effectiveness of Preschool Education International

Given the importance of sensitive periods early in life, an intervention at an early age seems to be an effective method to reduce developmental delays. Numerous researchers have argued that investing in early childhood yields higher returns at later ages (Howard-Jones, Washbrook, & Meadows, 2012; Shonkoff & Phillips, 2000). Heckman (2012) argued higher returns can be found in education, health and productivity outcomes. Particularly for the most disadvantaged groups, early intervention could counteract the buildup of developmental delays. Whether children really improve their skills because of early childhood education programs and whether it is an effective strategy for closing developmental gaps has been stirring global debate and research.

In general, investigating the potential impact of preschool education and relating it to a wide range of developmental outcomes is difficult, due to a number of practical and ethical restraints on research. For example, in real-life situations there is no random allocation of children into different educational interventions and there is an abundance of ‘relevant’ short term or long term cognitive and socioemotional outcomes. In this section, we will first highlight international findings on preschool education effectiveness, followed by insight into the Dutch situation specifically. Zooming in on the country level is interesting since generalization of outcomes and comparison of outcomes across countries is often hard because of essential

differences in structure and organization of the programs (De Haan, Leseman, & Elbers, 2011; Fukkink et al., 2015).

Burger (2010) reviewed key studies performed in Europe, North-America and Asia, focusing on the effectiveness of early childhood education in increasing cognitive performance in young children. He concluded that the clearest evidence for the potential efficacy of early childhood education stems from small-scale high-quality programs or model programs, using an experimental design. Well-known examples are the Perry Preschool Program (Schweinhart et al., 2005), the Abecedarian project (Barnett, 2008; Barnett & Masse, 2007) and HEAD start (Love et al., 2005; Puma et al., 2010), which all have yielded positive outcomes in children regarding diverse aspects of development, for example, cognitive outcomes and school readiness, labour market outcomes, income. However, these model programs demonstrate the potential effect of preschool education but are rather difficult to compare to the large-scale implementation of such programs. Therefore, Burger (2010) in his literature review on the effectiveness of early childhood education turned to medium- to largescale quasi-experimental studies, representing a more realistic picture of preschool education. He concluded that most studies found positive short-term effects, but much smaller long-term effects. Children from socio-economically disadvantaged families (the target group) tended to benefit slightly more than other children. But in general, early childhood education could not completely compensate for developmental disadvantages (Burger, 2010). Regarding the broader spectrum of development, such as socio-emotional aspects, a review study by Blok, Fukkink, Gebhardt, and Leseman (2005) showed that there are also no studies which have found significant effects on these outcomes. All in all, the general international agreement is that targeted preschool education programs work and that, due to the complexity of the issue, the effects are best studied in experimental research designs.

2.1.3 Early Childhood Education in the Netherlands

Organization and structure of early childhood education in the Netherlands.

The availability of and policies on early childhood education have been greatly extended nationally in the last decade (Meij, Mutsaers, & Pennings, 2009). Early childhood education (ECE or “Voor- en Vroegschoolse Educatie” (VVE) in Dutch) in the Netherlands is organized twofold. The first part of ECE takes place before primary school and is organized in child daycare centres and toddler playgroups or preschools and consists of at least 10 hours of active

program utilization in the groups (Nederlands Jeugdinstituut, n.d.-b). Daycare centres are generally open to children from 0 to 4 years of age and serve as a labour market instrument. Preschools or toddler playgroups, open to children from 2 or 2.5 years of age until the children attend primary school, specifically aim to stimulate child development. The second part of ECE, is organized in the first and second year of primary education when children are 4-5 and 5-6 years of age respectively. ECE therefore refers to a targeted educational program that is not provided in all child facilities by default. 'Group 3' or Grade 1 is the start of formal primary education, where no ECE is offered anymore.

The focus of this study lies on the first part of ECE that is provided in daycare centres and toddler playgroups, which we will refer to as Preschool Education Programs (PEP) in the remainder of this article. Though PEP had an initial focus on language skills, it currently also supports the development of number sense, motor and socio-emotional skills. Schools and municipalities receive monetary means from the government to aid children with a higher risk of developmental disadvantages (Nederlands Jeugdinstituut, n.d.-a). Initially, only the larger municipalities in the Southern region of the Netherlands received financial means to target at-risk children. In 2008 however, a financial impulse was provided to the Southern region of the Netherlands to ensure that even children outside the three largest municipalities in this region could benefit from the program (Rijksoverheid, 2008). Specific criteria have been formulated to identify target-group children whose parents are stimulated to send their child to a facility that offers PEP. Each Dutch municipality applies the same basic criterion concerning parental education level, which states that children of whom the parent(s) have at most an education level of the lower part of lower vocational secondary education, are 'at-risk' and parents are encouraged and financially supported to send their child to a facility that offers a preschool education program. The municipalities are free to add additional criteria, such as socio-emotional factors and language spoken at home but they receive no extra funding for this from the national government. Therefore, differences exist between municipalities in the extent to which monetary means are available and distributed for ECE purposes.

Effectiveness studies into early childhood education in the Netherlands.

Studies examining the effectiveness of preschool education programs influencing developmental outcomes show inconsistent results in the Netherlands (Driessen, 2004; Karssen, Van Der Veen, Veen, Van Daalen, & Roeleveld, 2013; Meij et al., 2009). Fukkink et

al. (2015) performed a meta-analysis investigating effectiveness studies and concluded that no significant influence of ECEC programs on child development outcomes, cognitive or socio-emotional, could be found. Research results were mostly moderated by the calendar year in which the study was performed (with less positive findings as time progresses), the correction for statistical background variables, and in general smaller effects were found if the program had already ended (so-called “fade-out” of the results). For example, Driessen (2004) did not find significant effects on child outcomes of preschool enrollment using a large-scale dataset from PRIMA (now continued under the name Pre-COOL) collected in the years 1996-2000. However, since the year 2000 investments in preschool education have increased. A more recent data collection that is often employed to study ECEC effectivity is the pre-COOL (and COOL) longitudinal cohort study. Bruggers, Driessen, and Gesthuizen (2014) and Karssen et al. (2013) for example used the Pre-COOL dataset to investigate the relation between preschool enrollment and/or a preschool education program and language and math outcomes, but did not find any positive significant results. Because of the longitudinal nature of the dataset there was a problem of attrition.

Leseman and Veen (2016) reported on Pre-COOL data and showed that target group children had a significant improvement of their previous lag in vocabulary and selective attention when they had been in a preschool education program, but work-play attitude improved less in target group children than in other children, creating a developmental gap. With regard to child outcomes that are measured with school tests, the authors show that for math test scores, children from their self-defined target groups lag behind and do not catch up. With regard to language, there is slightly more progression but the specific relation with ECEC could not be studied in depth yet. Although the study by Leseman and Veen (2016) provides interesting insights into the effects of ECEC on certain developmental aspects, it is not yet clear how it relates to the academic achievement measures schools employ, nor does the study shed light on what participation in preschool education adds (in comparison to not participating) for both target and non-target children. Akgunduz and Heijnen (2016) showed that an increase in the ECEC-budget in 2013 was associated with a smaller chance of remaining longer in kindergarten with 1 to 3 percentage points. However, doubling years in kindergarten is a complex phenomenon and the outcome of many different factors.

It is not surprising that mixed results have been found in the Netherlands, as many studies were constrained by methodological difficulties. First, it is difficult to realize a control

group let alone to assign children randomly to conditions. As stated earlier, participation rates of the target group children in the larger cities were 78% in 2014. Second, whether early childhood education is provided in childcare services and kindergarten or not, and for whom, varies per municipality and organization. This makes distinguishing between effects methodologically difficult (Leseman & Veen, 2016). Third, it requires a large sample size combined with a thorough baseline measurement and a long follow-up to reveal possible effects of the intervention. Fourth, choosing among possible outcome measures of child development or even choosing among potential aspects of ECEC influencing these outcomes is difficult. For instance, positive effects can arise due to different factors, such as having a sensitive responsive teacher (Tavecchio, 2008). Another example is the doubling of kindergarten years, used in a study mentioned earlier, in which the complex dynamics of doubling may be difficult to relate specifically to ECE. Fifth, there have been many recent policy and practical changes in the field which are highly likely to influence the outcomes of interest (VVE Platform, n.d.), but it is hard to control for this influence. For example, in 2010 a new law was introduced ("Wet OKE," n.d.), which was aimed at intensifying the ECEC focus and increasing quality. Policy changes are also related to the belief of several Dutch authors that the lack of strong findings is caused by differences in the extent of the implementation of the program, which causes too much variation (De Haan, Elbers, & Leseman, 2014; Tavecchio & Oostdam, 2013). Also, Dutch results may seem disappointing in comparison to international or US literature, due to a less intense nature of the preschool programs used in the Netherlands and a difference in population composition (Nap-Kolhoff et al., 2008).

Aims of the current study.

The current study provides new insights using a unique dataset. It comprises a common non-experimental, retrospective analysis technique on a large sample (Burger, 2010). Using recent data from a large, perennial data collection from the school years 11-12 to 14-15 (four cohorts), we analyse relations between preschool education program employed in the Netherlands before children enter kindergarten (before the age of 4) and their cognitive outcomes at the age of 5 to 6, irrespective of general preschool enrollment. No sampling was involved and also children who did not attend preschools were included, allowing comparison of existing groups in the population. Via the inclusion of a municipality variable, variations among municipalities, such as differing definitions of target groups, are accounted for. Since the measurement took place for multiple school years, the year of measurement is included as a control variable as well,

ensuring that policy changes on PEP are controlled for. Because of the financial impulse in the Southern region of the Netherlands, there is an increase of children enrolled in ECEC over the years, cancelling out negative selection into the ‘treatment’ group, thus making the sample more representative. Several variables related to SES are included in the dataset, such as parental education level, parental country of birth and language use at home between parents and children. Thus, the dataset allows us to highlight the target group as specified by the Dutch government (based on parental education level), while taking into account municipality- and school year-driven variation and relevant background variables.

This paper has two aims. First, we aim to explore the relationship between receiving PEP and the educational outcomes of children at age 5 to 6. Second, we aim to compare the impact of PEP for children of target groups and children of non-target groups. Based on previous findings, we expect that PEP has a positive relationship with children’s educational outcomes at a later age. We also expect an overall positive trend for all children, but a stronger relationship for target-group children than non-target-group children.

2.2 Method

2.2.1 Data

Data collection.

The data of Onderwijs Monitor Limburg stem from a large cooperative project between Maastricht University and schools, schools boards and local government. This data collection particularly aims to collect and analyze information about the educational development of students in order to foster educational improvement. The data collection was co-initiated by local school boards, ensuring that more than 95% of primary schools in the region participated. The data covered students’ school career starting in kindergarten, up to secondary school. Test scores were processed by a centre for data and information management to create an anonymized dataset for research purposes. For the current project, we utilize data from children in the last year of kindergarten (“group 2”) for the school years 2011-2012, 2012-2013, 2013-2014 and 2014-2015. We combined data of a parental questionnaire, test scores of the standardized Dutch CITO test of children aged 5-6 years old and school administrative data.

Administrative data.

School administrative data contain information concerning the child's date of birth and gender. The data also include information on the municipality in which the school is located. As stated earlier, each municipality uses different criteria to select target group children, with the exception of the basic criterion which every municipality uses (i.e. parental educational level). Therefore, the specific criteria determine whether parents of specific children in specific municipalities will be encouraged to utilize the program for their child. We select target group children in the sample based on the criterion of parental education level only and we account for any other differences in target group inclusion by including a municipality control variable. In total, children from 18 municipalities in the region are included in our dataset.

Parental questionnaire.

The parental questionnaire contains a range of questions regarding family and child characteristics. For our analysis, information about preschool use before the start of kindergarten, parental education level, language use at home and parental country of birth is extracted from the questionnaire. Parental education level was assessed based on the Dutch education system (Nuffic, 2015) and transformed into the International Standard Classification of Education (or ISCED) (Statistics, 2012) for analysis. There is no one-to-one correspondence between the ISCED classification and the PEP parental education data. The approach for this will be highlighted in the descriptive statistics section. Parents could indicate their country of birth in the questionnaire. This was then transformed into a variable indicating whether parents were born in the Netherlands or not. Parents born outside the Netherlands can be born in a Western country (such as Belgium or Germany) or a non-Western country (such as Turkey or Morocco), but since the variable will only be used as a control, these children are allocated to the same category. The variable indicating the language spoken at home consists of several levels of Dutch proficiency. We use questionnaires that are filled in by the mother or father of the child. We assume that questions regarding the spouse or partner refer to the other biological parent.

Test data.

Standardized language and mathematics tests (Cito, n.d.) are administered twice each school year in kindergarten; once halfway through the school year and once at the end of the school

year. Therefore there is a total of four test moments in kindergarten. The two tests administered in the first year are identical, as are the two tests administered in the second year. For this study, we used the first test administered in the second year of kindergarten to avoid test-retest effects that may occur at the end of the second year. Only test data administered in the months January or February and of children who had taken both the language and the math test were included. The Cito-test provides raw scores and so-called skill scores. The latter is a standardized version of the raw score. For the purpose of this paper, skill scores were used. Furthermore, the test data contain the date at which the Cito-test was administered, which allows exact calculation of the age of the child at the moment of the test. We added age in months as a control variable and restrict the sample so that only children aged 60 to 84 months were included in the sample – representing the normative age of children in the second year of kindergarten.

2.2.2 Sampling Procedure

For all four school years (2011-2012, 2012-2013, 2013-2014 and 2014-2015), a total of 19521 children were identified. Since the data are collected in the field, and the data collection is dependent upon several external factors such as participation of the school and parental questionnaire response, not all relevant information is available for all of the identified children. We include children from schools that provided test scores in all four years since there might be a certain bias as to why (and which) schools participate for a limited number of years. Furthermore, children for whom all relevant variables are present are included (i.e. information from the parental questionnaire, administrative data and test scores). We also only include children who were in group 2 once, meaning that children who doubled group 2 were excluded. This brings the dataset back to 4075 children.

2.2.3 Data Analyses

Analyses consisted of multiple regression analyses, in which a basic regression model is employed at first, which is extended in two models. We start with either the score on the language or math test as the outcome variable and participation in PEP as an independent variable of interest. In model 2, we add gender, age, school year and school as predictors as well as parental education level, parents' country of birth, language spoken at home and whether the child attended a daycare facility at all. In model 3, the municipality is added as a predictor variable, and the school variable is dropped.

2.3 Results

2.3.1 Descriptives of the Full Sample

The means and standard deviations of the skill scores on the language and math test can be found in Table 2.1. Within the sample, there was an almost equal division of gender across school years (52/48% M/F for 2011-2012, 52/48% M/F for 2012-2013, 50/50% M/F for 2013-2014, 50/50% M/F for 2014-2015). An independent samples t-test indicated that girls ($M=69.23$; $SD=10.17$) performed better on the language test than boys ($M=67.71$; $SD=9.81$), $t(4073) = -4.86$, $p < .01$, but that there was no significant difference for the math test between girls ($M=87.29$; $SD=11.19$) and boys ($M=88.20$; $SD=11$), $t(4073) = 2.60$, $p < .01$.

Table 2.1. Descriptive statistics of the CITO skill scores in the sample.

Skill score	Observations	Mean	Standard deviation	Min	Max
Language	4075	68.46	10.01	35	108
Math	4075	87.75	11.10	58	137

Age in months at the moment of the test was restricted from 60 to 84, representing the normative age of children in this group, with a stable mean across school years (for language/math 69.65/69.76 in 2011-2012, 69.27/69.32 in 2012-2013, 69.01/69.03 in 2013-2014 and 68.96/68.98 in 2014-2015). The correlation between the age in months and the test score in this sample appears to be small but significant (for language $r(4073) = .09$, $p < .001$ and for math $r(4073) = .13$, $p < .001$).

As for the different control variables used in the main analysis, means and standard deviations of the skill scores on the language and math test for different groupings of children can be found in table 2.2. The results show that children of low educated parents scored lower than children of highly educated parents, $F(4,4070) = 37.32$, $p < .01$ for language and $F(4,4070) = 31.78$, $p < .01$ for math. Table 2.3.1 provides a more detailed overview of the relationship between PEP and parental education level within the sample. This table shows that children who were receiving PEP were a mix of parental education levels. Therefore, analysis was first performed on the entire sample of children exploring the influence of PEP, followed by an analysis specifically on target group children, rather than only focusing on the target group children only.

Table 2.2. Overview of relevant variables in the sample.

Control variables	N(%)		Skill scores			
			Language		Math	
			Mean	SD	Mean	SD
Parental education						
Primary	30	(0.74)	62.03	11.71	79.27	7.62
Lower secondary	302	(7.41)	65.22	8.94	83.89	10.27
Upper secondary	1411	(34.63)	66.72	9.45	86.31	10.71
Bachelor	1684	(41.33)	69.84	9.88	88.95	11.14
Master	648	(15.90)	70.44	10.81	89.96	11.34
Total	4075					
Parental country of birth						
Both in NL	3558	(87.31)	68.87	9.98	88.11	11.15
One in NL	379	(9.30)	67.15	9.96	86.71	10.46
Both outside NL	138	(3.39)	61.44	7.82	81.36	9.38
Total	4075					
Language at home						
Dutch with both parents	3783	(92.83)	68.79	10.02	88.03	11.08
Dutch with one parent	259	(6.36)	64.91	8.78	84.34	10.85
Only a foreign language with both parents	33	(0.81)	57.88	7.63	82.24	9.96
Total	4075					
Childcare facility						
No attendance	334	(8.20)	67.55	10.23	86.85	12.21
Attendance	3741	(91.80)	68.54	9.99	87.83	11
Total	4075					
PEP						
No PEP	2589	(63.53)	68.21	10.07	87.44	11.17
PEP	1486	(36.47)	68.88	9.90	88.30	10.98
Total	4075					

Table 2.3.1 Overview of PEP use for different levels of parental education, only children who attended a childcare facility are included.

	Parental education level					
	Primary N (%)	Lower secondary N (%)	Upper secondary N (%)	Bachelor N (%)	Master N (%)	Total N (%)
no PEP	23 (76.67)	196 (64.90)	900 (63.78)	1055 (62.65)	415 (64.04)	2589 (63.53)
PEP	7 (23.33)	106 (35.10)	511 (36.22)	629 (37.35)	233 (35.96)	1486 (36.47)
Total	30 (100)	302 (100)	1411 (100)	1684 (100)	648 (100)	4075 (100)

Given the financial impulse provided to the region during the last decennium to increase PEP, we expected to see an increase of children receiving PEP. Table 2.3.2 shows that for target group children, as well as non-target group children, there was a rise in PEP usage per school year (from 26.09% to 33.33% and from 37.38% to 38.76% respectively). Therefore, we have therefore reason to assume that throughout the years, there was a less strong negative selection of children who participated in the program.

Table 2.3.2 Overview of PEP use for target group and non-target group children throughout school years.

School year	Target group			Non-target group		
	No PEP N(%)	PEP N(%)	Total N(%)	No PEP N(%)	PEP N(%)	Total N(%)
2011-2012	17 (73.91)	6 (26.09)	23 (100)	583 (62.62)	348 (37.38)	931 (100)
2012-2013	23 (69.70)	10 (30.30)	33 (100)	597 (66.78)	297 (33.22)	894 (100)
2013-2014	22 (66.67)	11 (33.33)	33 (100)	632 (63.39)	365 (36.61)	997 (100)
2014-2015	26 (66.67)	13 (33.33)	39 (100)	689 (61.24)	436 (38.76)	1125 (100)
Total	88 (68.75)	40 (31.25)	128 (100)	2501 (63.36)	1446 (36.64)	3947 (100)

2.3.2 Descriptives of target group sample

Target group children can be identified via various criteria. For the purpose of the current analysis, we defined target group children based on the basic criterium that applies to every municipality. Therefore, children who fit the criterion based on parental education level fit either in the primary education level or the lowest track of lower secondary education. The ISCED classification did not translate directly to the criterion. In total, 128 children – 30 of whom the parents had a primary education level and 98 of whom the parents had the lowest track of lower secondary education – fit the criterion of parental education level to be part of the target group, and less than one third of this group actually received PEP within the sample

(40 children). Therefore, we used parental education level to define the target group in the sample. Other relevant variables, such as language spoken at home and whether parents are born in the Netherlands served as control variables.

Within the sample, 334 children had at least one parent who was born in a foreign country. This country was either a Western or non-Western country. Also, 8.20% of the children within the sample did not attend any facility before they entered kindergarten (2.5% did not use any form of childcare, 5.69% used other forms of childcare such as family members or a childminder).

2.3.3 Multiple Regression Analysis

We regressed skill scores of the math and language Cito-test, on several relevant predictor variables (see Table 2.4) and on a dummy variable that indicated whether or not children received PEP. Gender was coded as a categorical variable with boys as the first and girls as the second category. Use of daycare was a dummy variable in which 1 meant that parents used child daycare. Table 2.4 shows the results of using this multiple regression analysis on the entire sample of 4075 children (including both the target and non-target group). Columns (1) display the results of the analyses without control variables, columns (2) with gender, age in months, school year, municipality, parental education level, parental country of birth and language spoken at home, whether the child attended a daycare facility, and which school the child attended as control variables. In columns (3) school is replaced with municipality as a control variable. See Table A in appendix A for an extended table.

Columns (1) show a significant relationship between preschool education and cognitive outcomes ($F(1, 4073) = 4.24, p < .05$ for language and $F(1, 4073) = 5.71, p < .05$ for math). Columns (2) demonstrate a significant positive relationship between preschool education and language test outcomes ($F(134, 3940) = 6.04, p < .001$) and math test outcomes ($F(134, 3940) = 5.39, p < .001$). Columns (3) depict a significant positive relationship for language test outcomes ($F(31,4043) = 13.52, p < .001$) and math test outcomes ($F(31,4043) = 12.11, p < .001$). Looking at the coefficient for PEP, which is the main interest of the analysis, we see a significant coefficient for both language and math in model 1. A coefficient of 0.67 for language can be interpreted as a rise in test scores of 0.67 points if the student received PEP before the age of 4.

Table 2.4 Overview of multiple regressions analysis (N = 4075) on how well preschool education predicts test outcomes.

OLS results of all children in the sample						
	Language (1)	Math (1)	Language (2)	Math (2)	Language (3)	Math (3)
PEP	0.67** (0.33)	0.86** (0.36)	0.65** (0.33)	1.08*** (0.36)	0.54* (0.33)	0.77** (0.36)
Gender			1.89*** (0.30)	-0.45 (0.33)	1.80*** (0.30)	-0.49 (0.34)
Age			0.35*** (0.04)	0.46*** (0.04)	0.31*** (0.04)	0.43*** (0.04)
Daycare use			0.56 (0.56)	0.46 (0.62)	0.07 (0.57)	-0.02 (0.63)
Constant	68.21*** (0.20)	87.44*** (0.22)	37.75*** (3.28)	49.03*** (3.67)	42.47*** (3.40)	50.89*** (3.79)
R Squared	.00	.00	.17	.16	.09	.09

Note. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 2.5.1 shows the results of multiple regression analysis for target group children, compared to non-target group children, specifically for outcomes in the language domain. See Table B.1 in appendix A for extended table. Columns (4) provide the results for the application of basic model (1) with no added predictor variables for target group (4a) and non-target group children (4b) separately (respectively, $F(1,126) = 3.26$, $p < .1$ for the target group and $F(1,3945) = 2.88$, $p < .1$ for non-the target group). In columns (5a) and (5b), model (3) was applied to language outcomes to allow comparison between target group and non-target group children, hereby including relevant predictor variables ($F(25,102) = 2.28$, $p < .01$ and $F(27,3919) = 8.56$, $p < .001$, respectively). The coefficient of PEP remained significant on a .1 level in columns (4a) and (4b) for both groups. The size of the coefficient was much larger for target group children as compared to non-target group children. In columns (5a) and (5b), the coefficients were no longer significant.

In table 2.5.2 the same analysis technique as in table 5.1 are presented for math outcomes. See Table B.2 in appendix A for extended table. Again in columns (6a) (6b), a basic

model (1) was employed and no predictors were added for target group and non-target group children ($F(1,126) = 3.36, p < .1$ and $F(1,3945) = 4.19, p < .01$, respectively). In columns (7a) and (7b) the same analysis strategy was employed as in model (3) for target group and non-target group children for math outcomes ($F(25,102) = 1.13, p > .1$ and $F(27,3919) = 8.22, p < .001$, respectively). The coefficients for PEP were significant on a .1 in column (6a) for target group children and on a .5 level in columns (6b) for non-target group children, but similar to the analysis for language, the size of the coefficient was much larger for the target group

Table 2.5.1 Overview of multiple regressions analysis on how well preschool education predicts language test outcomes.

OLS results of target (TG) and non-target (non-TG) group children for language				
	Language	Language	Language	Language
	TG (4a)	non-TG (4b)	TG (5a)	non-TG (5b)
PEP	3.34* (1.85)	0.56* (0.33)	1.46 (1.94)	0.44 (0.34)
Gender			1.64 (1.67)	1.74*** (0.31)
Age			0.50** (0.19)	0.25*** (0.04)
Daycare use			3.10 (2.57)	0.40 (0.59)
Constant	63.26*** (1.03)	68.39*** (0.20)	32.75** (15.08)	50.31*** (2.90)
N	128	3947	128	3947
R Squared	.03	.00	.36	.06
All control variables	no	no	yes	yes

Note. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Target group children are defined according to parental education level criteria.

Table 2.5.2 Overview of multiple regressions analysis to investigate how well preschool education predicts math test outcomes.

OLS results of target (TG) and non-target (non-TG) group children for math				
	Math	Math	Math	Math
	TG (6a)	non-TG (6b)	TG (7a)	non-TG (7b)
PEP	3.46* (1.89)	0.75** (0.37)	0.85 (2.20)	0.68* (0.37)
Gender			-1.27 (1.89)	-0.57 (0.35)
Age			0.48** (0.21)	0.38*** (0.04)
Daycare use			1.39 (2.91)	0.36 (0.66)
Constant	81.97*** (1.06)	87.63*** (0.22)	49.60*** (17.00)	61.00 (3.24)
N	128	3947	128	3947
R Squared	.03	.00	.22	.05
All control variables	no	no	yes	yes

Note. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Target group children are defined according to parental education level criteria.

children. When adding all relevant predictor variables, the coefficient of PEP remained significant at a .1 level for non-target group children. Table 2.6 shows Cohen's d effect sizes for all models employed. A value of .2 is considered to be small, .5 is seen as medium and .8 or above is large (Cohen, 1988).

2.3.4 Conclusion of results

The results indicate a significant positive relationship between PEP provided to children before the age of 4 and their cognitive outcomes, as measured by standardized school tests

Table 2.6. Overview of effect size Cohen's *d* for each regression model, specified for outcomes on language and math Cito-test.

Model	Language Cohen's <i>d</i>		Math Cohen's <i>d</i>
1	0.07		0.08
2	0.06		0.10
3	0.05		0.07
	a	b	
4	0.33	0.06	-
5	0.15	0.04	-
		a	b
6	-	0.31	0.07
7	-	0.08	0.06

(so-called Cito-test skill scores) at the age of 5-6 years old. In a basic regression model, including no control variables, the results indicated that when children received PEP, they scored 0.67 skill points higher on the language test and 0.86 skill points higher on the math test. In model 2 and 3, we included relevant control variables, namely parental education level, language spoken at home, parental country of birth, school year, gender, age, and whether the children attended a daycare centre or not. In model 2, the school of the child was included and in model 3 the municipality in which the school lies was included instead of the school. The results remained significant, indicating an increase in language skill scores of 0.65 points and a highly significant increase in math skill scores of 1.08 points in model 2. When running the same model, using the variable on municipality level instead of school level, the relationship remained significant indicating an increase of 0.77 skill points for math skills, but the results for language, a coefficient of 0.54, were not significant anymore at a .05 level. In models 4-7, we compared the results for the target group and non-target group children. We used municipality as an added control variable and not schools, since the latter had too many categories to use in this comparison. In model 4 and 6, where no extra predictors were added,

we saw mostly barely significant results for both language and math outcomes. However, the size of the coefficient is a lot larger for target group children, which is indicative for larger gains in test scores than the non-target group children. More specifically, target group children gained 3.34 skill points for language and 3.46 skill points for math if they received PE, whereas non-target group children gained 0.56 skill points for language and 0.75 skill points for math if they received PEP. In models 5 and 7, most of the results are not significant anymore at a .05 level. In general, the use of daycare did not contribute significantly to any of the models. Effect sizes were generally low, except for target group children when no controls are added, which are just above 0.3. A general rule of thumb is that effect sizes around .2 are considered small and around .5 medium. Provided with the sample size of this study, it may be concluded that a satisfactory effect size was found.

2.4 Discussion

Different experiences during early childhood have a great impact on brain development (Brown & Jernigan, 2012). Studies have shown that children from lower socioeconomic status (SES) backgrounds experience a higher risk of a developmental delay, even before they enter formal primary education at the age of 6 (Shonkoff & Phillips, 2000). These research findings have had a great impact on policymakers and many investments have been made in the development of early childhood enrichment via early childhood education (ECE) programs. In the Netherlands, ECE is organized both before and during primary education. The focus of this paper is on effectiveness of preschool education programs (PEP).

A basic criterion has been defined urging children of parents who are low educated to participate in PEP. In addition, municipalities can choose to add more criteria to be able to include more children, such as children with socio-emotional or language problems. Nonetheless, descriptive statistics revealed that, contrary to what is possibly the general expectation, children partaking in PEP in our sample were a mix of the population with respect to parental education level. This means that teachers had both target group and non-target group children in their groups. Therefore, children who did not have a higher risk of developmental disadvantage could still receive PEP. Such variety in receivers of PEP was also discussed by Borghans and Coppens (2015). The authors stated that children who do not attend any daycare facility may come from various backgrounds, such as families where mothers work fulltime, part-time or not at all. We also noticed a rise in the percentage of children enrolled in PEP in

our sample over the years, which is in line with policy changes aimed at increasing ECE participation of children. These descriptive statistics are therefore in line with the financial impulse provided to the region. The analysis was divided into two parts; with multiple regression analysis on the entire sample of 4075 children as well as analyses of target-group versus non-target group.

Our findings indicate that overall there was a positive relationship between PEP before the age of 4 and achievement test scores at the age of 5-6 when studying the entire sample of children. Results were significant for language and math when no predictor variables were added, but also when relevant predictors were added to the model, namely gender, age, language spoken at home, parental country of birth, parental education level, school year and school of the child. We see the highest effect sizes around .3 for target group children in a basic regression model, which corresponds to a small to medium effect size (Cohen, 1988). Even when adding all relevant control variables, the effect sizes for target group children remained higher than for non-target group children. Provided with our field of study, even such effect sizes that would normally be categorized as small, can be considered highly interesting when keeping the context of educational outcomes in mind (Durlak, 2009).

When the predictor variable on school level is replaced by with the municipality of the school, results for language lose some significance, whilst for math, similar results were found. When splitting the sample for target group and non-target group children, the gain is a lot higher for children in the target groups when no predictor variables are added. When adding the relevant predictor variables and municipality, results are overall of low impact/effect size/power or non-significant. The use of daycare does not contribute significantly to any of the models. This means that whether children attend a daycare facility or toddler playgroup, this attendance solely does not contribute significantly to cognitive outcomes later on. Our findings are in contrast with the study performed by Nap-Kolhoff et al. (2008) who concluded that differences between children's scores in the short term (group 2) could be explained by variables such as gender and SES.

The overall positive trend in our results is in accordance with the conclusions of Currie (2001), who stated that preschool education can be effective for many more children than only the disadvantaged group. We found a significant positive relationship between PEP and language and math skills, which contrasts the findings reported by Leseman and Veen (2016).

The authors found a positive relationship of PEP with language aspects (i.e. vocabulary) and no relationship with math skills while controlling for relatively similar control variables. However, the authors looked at whether the gap between target and non-target children became smaller while following PEP (i.e. whether they ‘catch up’), whereas in our study, we examined whether children who follow a preschool education program do better at a later age than children who did not follow a preschool education program. We can also explain this difference in findings by a difference in sample composition or the use of a baseline measurement by Leseman and Veen (2016). The use of a baseline measurement offers more comparative elements in the dataset. By including the municipality control variable in our analysis, we also overcome some problems of differences between municipalities which was sometimes used as an explanation for the lack of positive findings in other studies (Bruggers et al., 2014). Our findings contrast the conclusion in Fukkink et al. (2015) which entailed no effect of ECEC in the Netherlands. Burger (2010) has also stated that many quasi-experimental studies using retrospective designs suffer from not being able to separate the effects of preschool enrollment and a targeted program. In our sample, we include information in the models both on daycare use and whether the child belongs to a target group and we saw that the role of daycare attendance in the models is not significant.

Our study has certain limitations. Working with parental questionnaires entails that our dataset is partly self-reported and potentially biased by memory mistakes, false answers or a selection bias as to which parents fill in the questionnaire. An example is the small number of lowly educated parents who filled in the questionnaire. However, Golsteyn & Hirsch (2019) analyzed the same dataset for group 8 children and concluded that non-response from schools is generally more severe than parental non-response based on educational level. The latter we controlled for by only including schools that provided student test scores for all four cohorts. Therefore, some schools are excluded from the data. We believe these schools have potentially interesting profiles since they decided not to participate each school year for reasons unknown. Results could potentially even be enhanced if such schools were included. This is, however, mere speculation. The results show that, after adding all relevant control variables, the relationship between PEP and cognitive outcomes appears to be present in this subsample to some extent. Therefore, we speculate that we captured a glimpse of the relationship between PEP and cognitive outcomes for the whole target group in the population. True relationships may even be stronger than found in the current study due to the school bias.

Another limitation is that there is no baseline measure to assess cognitive function when the children started in PEP. We expect that this would yield more insight into the individual growth of children, but only for those specific tasks, and probably only for children who attend a daycare facility. Furthermore, there was no control group that allows a clear comparison between students. Provided with the high difficulty of performing research in this field, we probably reach a limit of what we can prove without performing an actual randomized controlled experiment. The biggest disadvantage of our study is that we did not take into account the quality, intensity or length of the PEP intervention. Choices regarding variable inclusion have to be made to avoid overly extensive inclusion of detailed variables, creating a multitude of subgroups, such as number of children in each group, type of program, duration of program, intensity, etc. In general, these characteristics often determine program quality and even process quality, which has been shown to be an important aspect of ECEC effectiveness (Fukkink et al., 2015; Inspectie van het Onderwijs, 2016; Leseman & Veen, 2016; Van Huizen & Plantenga, 2015). Last, it is possible that the result we found fade-out in the long run, which has been reported by other authors (Barnett, 2011). Nevertheless, this effect has not been confirmed (Van Huizen & Plantenga, 2015).

2.5 Conclusion

Our study finds a significant positive relationship between PEP provided to children before they enter kindergarten and their achievement on school tests at the age of 5-6, for language as well as math skills. When adding predictors to the model for municipalities, schools, years, and thus for policy changes over time, differences between regions, as well as for use of daycare facilities and a preschool education program, background factors such as socioeconomic status, language spoken at home, and parental country of birth, a positive relationship between PEP and higher educational outcomes largely remains for the entire sample of children. For target group children particularly, gains are particularly high when a basic model is employed but fades when more predictors are added, which can be due to a smaller sample size of the target group children. The effect sizes for the general sample are small, but provided in the context of educational outcomes, these can be very significant and relevant to understand, since particularly for target group children the highest effect sizes were found.

2.6 References

- Akgunduz, E., & Heijnen, S. (2016). *Impact of funding targeted pre-school interventions on school readiness: Evidence from the Netherlands*. Retrieved from <https://www.cpb.nl/sites/default/files/publicaties/download/cpb-discussion-paper-328-impact-funding-targeted-pre-school-interventions-school-readiness.pdf>
- Barnett, W. S. (2008). *Preschool education and its lasting effects: Research and policy implications*. Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved from <http://epicpolicy.org/publication/preschooleducation>
- Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science*, 333(6045), 975-978.
- Barnett, W. S., & Masse, L. N. (2007). Comparative benefit–cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review*, 26(1), 113-125.
- Blok, H., Fukkink, R. G., Gebhardt, E. C., & Leseman, P. P. (2005). The relevance of delivery mode and other programme characteristics for the effectiveness of early childhood intervention. *International Journal of Behavioral Development*, 29(1), 35-47.
- Borghans, L., & Coppens, K. M. (2015). Kind of arbeidsmarkt? *Economisch Statistische Berichten*, 100(4718), 560-560.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, 53(1), 371-399.
- Brown, T. T., & Jernigan, T. L. (2012). Brain development during the preschool years. *Neuropsychology Review*, 22(4), 313-333. doi:10.1007/s11065-012-9214-1
- Bruggers, I., Driessen, G., & Gesthuizen, M. (2014). Voor-en vroegschoolse voorzieningen, effectief of niet?-De samenhang tussen deelname aan voor-en vroegschoolse voorzieningen en de taal-en rekenprestaties van leerlingen op de korte en langere termijn. *Mens en maatschappij*, 89(2), 117-150.
- Burger, K. (2010). How does early childhood care and education affect cognitive development? An international review of the effects of early interventions for children from different social backgrounds. *Early Childhood Research Quarterly*, 25(2), 140-165. doi:10.1016/j.ecresq.2009.11.001

- Cito. (n.d.). *LVS-toetsen primair en speciaal onderwijs*. Retrieved from http://www.cito.nl/onderwijs/primair%20onderwijs/lvs_toetsen
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Currie, J. (2001). Early childhood education programs. *The Journal of Economic Perspectives*, 15(2), 213-238.
- De Haan, A. K., Elbers, E., & Leseman, P. P. (2014). Teacher-and child-managed academic activities in preschool and kindergarten and their influence on children's gains in emergent academic skills. *Journal of Research in Childhood Education*, 28(1), 43-58.
- De Haan, A. K., Leseman, P. P., & Elbers, E. (2011). *Pilot gemengde groepen 2007-2010. Onderzoeksrapportage oktober 2011*. Retrieved from <https://www.spelenderwijsutrecht.nl/Portals/0/DocumentenExtern/Pilot%20gemengde%20groepen%20eindrapport.pdf>
- Driessen, G. W. J. M. (2004). A large-scale longitudinal study of the utilization and effects of early childhood education and care in The Netherlands. *Early Child Development and Care*, 174(7-8), 667-689. doi:10.1080/0300443042000187158
- Durlak, J.A. (2009). How to Select, Calculate, and Interpret Effect Sizes. *Journal of Pediatric Psychology*, 34(9), 917-928.
- Fox, S. E., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development*, 81(1), 28-40.
- Fukkink, R., Jilink, L., & Oostdam, R. (2015). Met een blik op de toekomst: een meta-analyse van de effecten van VVE op de ontwikkeling van kinderen in Nederland. Retrieved from https://www.researchgate.net/publication/283944720_Met_een_blik_op_de_toekomst_Een_meta-analyse_van_de_effecten_van_VVE_op_de_ontwikkeling_van_kinderen_in_Nederland
- Fukkink, R., Jilink, L., & Oostdam, R. (2017). A meta-analysis of the impact of early childhood interventions on the development of children in the Netherlands: an inconvenient truth? *European Early Childhood Education Research Journal*, 1-11.
- Golsteyn, B.H.H., & Hirsch, S. (2019). Are estimates of intergenerational mobility biased by non-response? Evidence from the Netherlands. *Social Choice and Welfare*, 52, 29-63. doi: <https://doi.org/10.1007/s00355-018-1138-0>

- Heckman, J. J. (2012). *Invest in early childhood development: Reduce deficits, strengthen the economy*. Retrieved from <https://www.co.monterey.ca.us/home/showdocument?id=17884>
- Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P., & Yavitz, A. (2010). *A new cost-benefit and rate of return analysis for the Perry Preschool Program: A summary*. Retrieved from http://jenni.uchicago.edu/papers/Heckman_Moon_etal_2010_NBER_wp16180.pdf
- Hertzman, C., & Boyce, T. (2010). How experience gets under the skin to create gradients in developmental health. *Annual review of public health, 31*, 329-347.
- Howard-Jones, P. A., Washbrook, E. V., & Meadows, S. (2012). The timing of educational investment: A neuroscientific perspective. *Developmental Cognitive Neuroscience, 2*, S18-S29. doi:10.1016/j.dcn.2011.11.002
- Inspectie van het Onderwijs. (2016). *Investeren loont! Eindrapportage monitor kwaliteit voor- en vroegschoolse educatie in de 37 grote steden in 2015/2016*. Retrieved from <https://www.onderwijsinspectie.nl/documenten/rapporten/2017/02/13/eindmonitor-kwaliteit-voor--en-vroegschoolse-educatie-g37-2015-2016>
- Karssen, A., Van Der Veen, I., Veen, A., Van Daalen, M., & Roeleveld, J. (2013). *Effecten van deelname aan en kwaliteit van voor-en vroegschoolse educatie op de ontwikkeling van kinderen*. Amsterdam: Kohnstamm Instituut.
- Kautz, T., Heckman, J. J., Diris, R., Ter Weel, B., & Borghans, L. (2014). *Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. Retrieved from <https://www.nber.org/papers/w20749.pdf>
- Knudsen, E. I. (2004). Sensitive Periods in the Development of the Brain and Behavior. *Journal of Cognitive Neuroscience, 16*(8), 1412-1425. doi:10.1162/0898929042304796
- Knudsen, E. I., Heckman, J. J., Cameron, J. L., & Shonkoff, J. P. (2006). Economic, neurobiological, and behavioral perspectives on building America's future workforce. *Proceedings of the National Academy of Sciences, 103*(27), 10155-10162.
- Leseman, P. P., & Veen, A. (2016). *Ontwikkeling van kinderen en relatie met kwaliteit van voorschoolse instellingen. Resultaten uit het pre-COOL cohortonderzoek. (947)*. Retrieved from Amsterdam: <http://kohnstammstituut.uva.nl/ki947.html>
- Love, J. M., Kisker, E. E., Ross, C., Raikes, H., Constantine, J., Boller, K., . . . Vogel, C. (2005). The Effectiveness of Early Head Start for 3-Year-Old Children and Their

Parents: Lessons for Policy and Programs. *Developmental Psychology*, 41(6), 885-901. doi:10.1037/0012-1649.41.6.885

- Meij, H., Mutsaers, K., & Pennings, T. (2009). *Effectiviteit van voor-en vroegschoolse programma's in Nederland*. Retrieved from <http://www.nji.nl/publicaties/EffectiviteitVVEinNederland.pdf>.
- Nap-Kolhoff, E., van Schilt-Mol, T., Simons, M., Sontag, L., van Steensel, R., & Vallen, T. (2008). *VVE onder de loep: een studie naar de uitvoering en effectiviteit van voor-en vroegschoolse educatieve programma's*. Retrieved from <https://www.nro.nl/wp-content/uploads/2014/11/VVE-onder-de-loep.pdf>
- Nederlands Jeugdinstituut. (n.d.-a). Financiering. Retrieved from [https://www.nji.nl/nl/Kennis/Dossier/Voor-en-vroegschoolse-educatie-\(vve\)/Financiering](https://www.nji.nl/nl/Kennis/Dossier/Voor-en-vroegschoolse-educatie-(vve)/Financiering)
- Nederlands Jeugdinstituut. (n.d.-b). Voorzieningen met voor- en vroegschoolse educatie. Retrieved from <https://www.nji.nl/nl/Kennis/Dossier/Voorzieningen-met-voor--en-vroegschoolse-educatie>
- Noble, K. G., Tottenham, N., & Casey, B. J. (2005). Neuroscience Perspectives on Disparities in School Readiness and Cognitive Achievement. *The Future of Children*, 15(1), 71-89. doi:10.1353/foc.2005.0006
- Nuffic. (2015). *Education system The Netherlands* (version 4, January 2015). Retrieved from Retrieved from Nuffic website: <https://www.nuffic.nl/en/publications/find-a-publication/education-system-the-netherlands.pdf>
- Puma, M., Bell, S., Cook, R., Heid, C., Shapiro, G., Broene, P., . . . Spier, E. (2010). *Head Start Impact Study. Final Report*. Washington, DC: U.S. Department of Health and Human Services, Administration for Children and Families.
- Rijksoverheid. (2008). *Honderd procent dekking VVE in Zuid-Limburg*. Retrieved from <https://www.rijksoverheid.nl/actueel/nieuws/2008/09/29/honderd-procent-dekking-vve-in-zuid-limburg>
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: the High/Scope Perry Preschool study through age 40*. Retrieved from http://nieer.org/wp-content/uploads/2014/09/specialsummary_rev2011_02_2.pdf
- Shonkoff, J. P. (2007). *The timing and quality of early experiences combine to shape brain architecture* (Working paper no. 5). Cambridge, MA: National Scientific Council on the Developing Child.

- Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighborhoods: The science of early childhood development*: National Academies Press.
- UNESCO Institute for Statistics (2012). *International Standard Classification of Education: ISCED 2011*. Montreal, Quebec: UNESCO Institute for Statistics.
- Tavecchio, L. (2008). Succesfactoren in programma's op het gebied van voor-en vroegschoolse educatie. *Kind en Adolescent Review*, 15(3), 353-365.
- Tavecchio, L., & Oostdam, R. (2013). De invloed van onderliggende factoren van schoolrijpheid in relatie tot de voor-en vroegschoolse educatie. *Pedagogiek*, 33(1), 37-48.
- UNESCO (2015) *EFA Global Monitoring Report: Education for All 2000–2015: Achievements and Challenges*. Paris: UNESCO.
- Van Huizen, T., & Plantenga, J. (2015). *Universal child care and children's outcomes: A meta-analysis of evidence from natural experiments* (Tjalling C. Koopmans Research Institute, Discussion Paper Series 15-13). Retrieved from Utrecht School of Economics website: <https://www.uu.nl/en/organisation/utrecht-university-school-of-economics-use/research/discussion-papers/discussion-papers-2015>
- VVE Platform. (n.d.). Wet OKE. Retrieved from <http://www.vveplatform.info/dossiers/wet-oke/>
- Ministerie van Onderwijs, Cultuur en Wetenschap. (2015). *Evaluatie Wet ontwikkelingskansen door kwaliteit en educatie (Wet OKE) en specifieke uitkering gemeentelijk onderwijsachterstandenbeleid*. Retrieved from <https://www.rijksoverheid.nl/binaries/rijksoverheid/documenten/kamerstukken/2015/06/29/kamerbrief-over-wet-oke/kamerbrief-over-wet-oke.pdf>



Chapter 3

Measurement of executive functioning using a playful robot in kindergarten

Urlings, C.C., Coppens, K.M., & Borghans L. (2019). Measurement of executive functioning using a playful robot in kindergarten. *Computers in the Schools*, 36(4), 255-273. doi: 10.1080/07380569.2019.1677436

3.1 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), which allow higher-order behavioural 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, i.e. kindergarteners, the question is raised whether robots can be used to overcome some of the difficulties of traditional EF testing in kindergarteners.

3.1.1 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 even in the classroom nowadays 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 onwards, which beforehand were neglected or difficult to teach. In essence, technological applications and robots are employed in education as a means of teaching, stimulation and learning and possibly measurements.

3.1.2 Understanding Executive Functioning in Kindergarteners

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

as young as three to four years of age (Hendry, Jones, & Charman, 2016). However, since children often do not possess sufficient vocabulary or attention span for most readily available tasks 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 play a crucial role in measuring EF. Furthermore, Chaytor and Schmitter-Edgecombe (2003) found that the ecological validity of neuropsychological tests, which are used to measure executive functions, memory and attention processes, is mediocre at best when predicting everyday cognitive functioning skills. The authors stated that though the extensive test batteries may yield useful insights in clinical populations, the benefits might be lower when trying to understand healthy populations. As more technological means have become available, this study will highlight the potential of using robots in relation to measurement issues in kindergarteners.

3.1.3 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 utilized in programming a robot and that 7-year olds 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 utilized, 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 education and being acquainted with inquiry-based learning. Benitti (2012) performed a literature review, investigating whether robots were utilized in education in a broader sense. The author found that 80% of the studies reviewed focused on physics and mathematics.

3.1.4 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 behaviour 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, along with the potential of high ecological validity.

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 seems 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 behaviour from children, potentially yielding measures with higher ecological validity.

3.1.5 The Bee-Bot

The Bee-Bot has been found to be a very suitable robot for kindergarteners. The toy can be programmed to navigate it 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 the implementation of 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 behaviour 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, investigating multiple solutions. Janka (2008) described that the Bee-bot mostly elicited enthusiastic behaviour in children, reinforcing the attractiveness for open-ended activities. The author 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), visuospatial skills and attention. The authors concluded that EF, more specifically working memory and inhibition, improved after the Bee-Bot intervention. Based on these studies, we conclude that the Bee-Bot is a robot that 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.

3.1.6 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 let kindergarteners perform a Bee-Bot task and systematically quantify their behaviour and outcomes. We relate these outcomes to the children's outcomes on several tests that measure EF. In order to gain sufficient insight into the relationship between outcomes on the Bee-Bot task and EF, we include 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 non-verbal intelligence measure. The scores on these tasks will be compared to observables from the Bee-Bot task, such as the time they need to think before acting (thinking time), the time needed to solve the task (execution

time) and the number of errors they make. The EF tasks included are 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. If we utilize the same rationale for solving the Bee-Bot task, we hypothesize 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 include.

In sum, in this study, we examine the link between performance on Bee-Bot tasks and performance on standardized EF tasks. Our main research question is: What EF are measured when children program a Bee-Bot? We will first determine whether gender and age influence the performance on both the Bee-Bot and the EF tasks. Next, we model the Bee-Bot data (thinking time, execution time and errors made) into a latent variable, which we then relate 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, flexibility. Since we aim to get a good understanding of the relation of EF, we will include a broad spectrum of EF tests.

3.2 Method

3.2.1 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 should 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 their performance in some way (e.g. epilepsy, autism).

Bee-Bot task.

We developed three tasks for the children to perform in increasing difficulty for testing (figure 3.1c). 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.

The researcher observed specific behaviours of the children whilst they were engaging with the Bee-Bot using a stopwatch and scoring form. The first variable registered was thinking time in seconds, for which measurement started after the instruction has been given and stopped when the child presses 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 until the Bee-Bot reached the end position. The third variable is the number of errors children make during each assignment. The following errors were registered: pressing a button too often or not often enough and pressing the wrong button. The protocol for the Bee-Bot assignments and the scoring sheet which were devised for this study can be found in Appendix B.

Raven 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 fits the pattern from 6 options depicted below on the same page. The booklet contains 36 items, divided into three sets (A, Ab and B). The RCPM is often used to estimate the general intelligence of children (Raven, Raven, & Court, 1998). Lezak, Howieson, Loring, and Fischer (2004) states the task also measures 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 following studies yielded satisfactory reliability when assessed by split-half or retest methods. The score is the total number of correct items (max. 36).

Tower of London (ToL).

The Tower of London task stems from the NEPSY test battery (Korkman, Kirk, & Kemp, 1998). Designed for children from 5 to 12 years old, the test consists of moving three balls on

three pegs in such a way 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. It measures planning abilities and problem-solving (Unterrainer et al., 2004). Studies with adults indicate also that the ToL leads 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 (FPT).

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 utilized a 1-minute restraint for this task. Data on reliability are sparse and mostly focus on adults and normal populations (Strauss, Sherman, & Spreen, 2006). Tucha, Aschenbrenner, Koerts, and Lange (2012) reported strong relations between performance on the FPT, processing speed and mental flexibility, as well as medium relations with figural short-term memory, problem-solving and inhibition. We utilized the total number of correctly drawn figures.

Verbal fluency task (VFT).

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 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-Solís, and Bernal (2006) showed VFT correlates with immediate and delayed verbal memory. We utilized the total number of correctly named animals.

(Reverse) Digit span (RDS/DS).

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 backwards. 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 task has been successfully utilized 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) reports that DS relies on attention, whereas RDS relies on mental tracking or working memory in adults.

Mazes (MA).

The Mazes 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 a maximum amount of errors made (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).

3.2.3 Procedure

Testing was divided into 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 the gender of the child. Tests were administered in the school, in a separate room to accommodate one-to-one testing.

3.2.4 Analyses

For our analyses, we use the following variables. We included gender and age to understand whether these may confound the relation between Bee-Bot and EF outcomes. Regarding Bee-Bot variables, we include for each of the three assignments thinking time, execution time and number of errors. For EF tasks we include 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 create a latent variable ‘Bee-Bot’, 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.

3.3 Results

3.3.1 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.

Table 3.1a and 3.1b shows descriptive statistics and frequencies for the Bee-Bot task and the EF test battery. Execution time increased as the level of difficulty of the tasks increases, 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, the range in the RDS was rather low. Therefore, only DS was utilized in the analysis further on. As can be seen in Table 3.2, 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 3.3 shows the correlations between the Bee-Bot measures and 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.

Table 3.1a Descriptive information on thinking time and execution time of the three Bee-Bot tasks.

Variables	N	Mean	Std. Dev.	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	8.84	49.73
Execution time 2	65	31.47	11.57	15.57	69.11
Execution time 3	65	53.42	18.02	26.37	115.95
ToL	65	12.89	1.79	8	16
DS	65	6.62	1.77	4	11
RDS	65	2.63	1.26	0	5
VFT ^a	65	10.94	3.77	3	21
RCPM	65	22.92	4.45	9	34
FPT ^a	65	6.79	2.36	1	11
MA	63	10.35	2.19	7	18

Note. ^aNumber of correct answers.

Table 3.1b Frequencies on the number 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 labelled as Errors 1, Errors 2 and Errors 3 in all following analysis, indicating the number of mistakes made by children per Bee-Bot assignment.

Table 3.2 Intercorrelations of the measures of the three Bee-Bot tasks.

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

Note. * $p < .1$, ** $p < .05$, *** $p < .01$. ^aThe abbreviations T1, T2 and T3 refer to Thinking time 1, Thinking time 2 and Thinking time 3. Ex1, Ex2 and Ex3 refer to Execution time 1, Execution time 2 and Execution time 3. Er1, Er2 and Er3 refer to Errors 1, Errors 2 and Errors 3.

Table 3.3 Correlations between the Bee-Bot task measures and the EF test battery

Bee-Bot measures	EF test battery outcomes					
	ToL	DS	FPT	MA	VFT	RCPM
Thinking1	-.12	-.14	.09	-.05	-.02	-.04
Thinking2	-.06	-.01	-.09	.03	-.07	-.24
Thinking3	.02	.01	-.04	-.03	-.12	-.17
Execute1	-.33***	-.19	-.01	-.01	-.10	-.28**
Execute2	-.16	-.06	-.07	.02	-.21*	.03
Execute3	-.20	-.08	-.05	-.02	-.13	.02
Errors1	-.23*	-.17	-.02	-.18	-.06	-.05
Errors2	-.43***	-.45***	-.04	-.17	-.29**	-.45***
Errors3	-.45***	-.36***	-.09	-.01	-.41***	-.21*

Note. * $p < .1$, ** $p < .05$, *** $p < .01$.

3.3.2 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 do not want to include the extent to which the EF tasks predict 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 freed and the latent variable was constrained to a mean of 0 and a variance of 1. We simultaneously regressed the latent variable ‘Bee-Bot’ 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 missings in modelling the data. The chi-square of the initial factor model was not significant at a .05 level, with χ^2 (90, N=65) = 110.84, $p > .05$. RMSEA was .06, the CFI/TLI values were 0.84/0.81. 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. = 0.47). The standardized factor loadings of the adjusted model can be found in table 3.4 and the standardized coefficients in table 3.5. 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, $r(63) = 0.47$, $p < .01$. Figure 3.2 shows a graphical depiction of the model. The fit of this model was good, with χ^2 (89, N=65)=95.16, $p > .1$, RMSEA = .03, CFI/TLI = 0.95/0.94. We also ran the model individually, in which the latent variable Bee-Bot predicts

Table 3.4 Standardized factor loadings of Bee-Bot variables on latent variable ‘Bee-Bot’.

Bee-Bot variables	Estimate	S.E.	Est./S.E.	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	0.29	.77
Execution time 3	.12	.14	0.88	.38
Errors 1	.26	.13	2.01	.05
Errors 2	.77	.07	10.52	.00
Errors 3	.73	.08	9.17	.00

Table 3.5 Standardized coefficients of regression of the EF test battery outcomes on latent variable ‘Bee-Bot’.

EF test outcomes	Estimate	S.E.	Est./S.E.	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	-0.89	.38
MA	-.16	.14	-1.15	.25

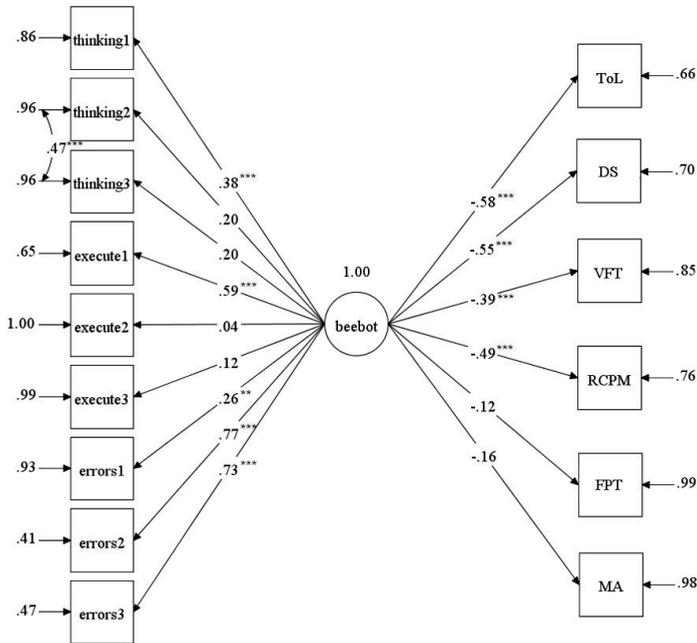


Figure 3.2 Graphical depiction of the model, including standardized estimates and residuals (***) $p < .01$, (**) $p < .05$, (*) $p < .1$).

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.

3.4 Discussion

Robots are used in education mostly to enhance learning in children, specifically related to programming, technology and engineering skills. In this study, we explore 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. Behaviours of the children were observed while they solved a task using the robot, in which the robot has 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).

3.4.1 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 variable 'Bee-Bot' 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 encompasses 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.

3.4.2 Implications of the findings

The results show with which EF the latent variable ‘Bee-Bot’ 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 labelled 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 behaviour 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, since aggregated performance on a Bee-Bot task in our study could significantly predict 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 same 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 relation 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 relationship 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 after 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 Bee-Bot latent 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 s/he 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 the Mazes task.

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 kindergarteners 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 are related to the children's performance as well.

3.4.3 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 future research in this area should include as well.

3.5 Conclusion

This study shows 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 and if 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 behaviour 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.

3.6 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-Solís, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of Semantic Verbal Fluency (ANIMALS). *International Journal of Psychology*, 41(5), 324-332.
- Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, 58(3), 978-988.
- 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.
- 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.
- 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.
- 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 New York.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, 72(4), 1032-1053.
- 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.
- 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 Intl. Conf. on Simulation, Modeling and Programming for Autonomous Robots. Venice, Italy.

- Di Lieto, M. C., Inguaggiato, E., Castro, E., Cecchi, F., Cioni, G., Dell’Omo, M., . . . Sgandurra, G. (2017). Educational robotics intervention on executive functions in preschool children: a pilot study. *Computers in Human Behavior*, *71*, 16-23.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168.
- 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. Toyoma, 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*. Proceedings of 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: Precursors, predictors and patterns. *Developmental Review*, *42*, 1-33.
- Highfield, K., Mulligan, J., & Hedberg, J. (2008). *Early mathematics learning through exploration with programmable toys*. In Figueras, O., Cortina, J.L., Alatorre, S., Rojano, T., & Sepúlveda, A.. 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.
- Janka, P. (2008). *Using a programmable toy at preschool age: Why and how*. Proceedings of Intl. Conf. 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.
- 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., Resing, W.C.M. & Verhaeghe, P. (2002). *WISC-III NL. Handleiding*. London: The Psychological Corporation.

- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. New York, US: Oxford University Press.
- Mitnik, R., Nussbaum, M., & Soto, A. (2008). An autonomous educational mobile robot mediator. *Autonomous Robots*, 25(4), 367-382.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, 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.
- Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory*, 7(2), 209-231.
- 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, Texas: 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.
- 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.
- 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 8-week robotics curriculum in pre-kindergarten through second grade. *International Journal of Technology and Design Education*, 26(1), 3-20.
- 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*, 12, 203-219.
- Taylor, R. (1990). Interpretation of the correlation coefficient: a basic review. *Journal of diagnostic medical sonography*, 6(1), 35-39.

- 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.
- 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 Experimental Neuropsychology*, 26(6), 846-856.
- 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.
- 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.
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33-35.



Chapter 4

A digital wooden tabletop maze for estimation of cognitive capabilities in children

Christopher, S.M., Urlings, C.C., Bongarth, van den H., Coppens, K.M., Hurks, P.P.M., Borghans, L., & Möckel, R. (2019). A Digital Wooden Tabletop Maze for Estimation of Cognitive Capabilities in Children. In: Salichs M. et al. (eds) Social Robotics. ICSR 2019. *Lecture Notes in Computer Science*, vol 11876. Springer, Cham.

4.1 Introduction

Cognitive assessments are useful to monitor a child's cognitive development (Blair & Razza, 2007; Borella, Carretti, & Pelegrina, 2010). The field of psychology knows a variety of standardized tests targeting a variety of cognitive aspects. The Porteus Maze test (Porteus, 1959) for instance is a nonverbal intelligence test developed to estimate the psychological planning capacity and inhibition of a participant (Kirsch et al., 2006). The Mazes subtest in WPPSI-R is another widely used test for assessing the cognitive development of young children (Kaufman & Lichtenberger, 2000). During these tests, children are confronted with mazes of varying complexity and instructed to solve as many mazes as possible within a given time. For this, children must trace through the mazes printed on paper with a pencil.

A key challenge with such standardized tests is that they rely on a professional observer to document the behaviour of the participating children. A detailed recording of all potentially relevant behavioural information becomes not only impossible but also sometimes unreliable (Carlson, 2005). For instance, to reach a high score in the Porteus Maze test, children are asked to solve mazes as efficiently as possible – by avoiding crossing maze lines, going into dead ends, without moving backwards, and lifting the pencil. The professional observer must record these features from the solved paper mazes by hand and calculate the final score. There might also be inconsistencies in the final score due to the complex scoring procedures, interrater reliability and administration difficulties (Kaufman, 1990). Other potentially valuable information, like for instance at which maze position and for how long a child stopped drawing because the child planned the next actions, gets lost. Another criticism of standardized tests is that they require structured environments and are typically used under well-controlled conditions. Standardized test thus can be stressful if the child has to leave his/her well-known environment or boring (Fleege, Charlesworth, Burts, & Hart, 1992). As a result, the child's motivation to participate in the test might drop and impact the test results (Borghans, Meijers, & Ter Weel, 2008).

To allow for maintaining high engagement of children during tests, to reduce the workload of testing professionals and to increase measurement accuracies, computer-based tests using game environments can present a valuable alternative to standardized pencil-and-paper tests (Pillay, 2002). Such computer-based tests have the additional advantage that they allow for collecting large amounts of detailed data in real-time. However, computer-based tests

also always test a child's ability to adapt to and to handle the computer environment. In addition, computer-based tests cannot fully reassemble the interaction with and perception of natural environments. As a result, computer-based tests are often not suitable for young children for instance because of underdeveloped fine motor skills.

A better alternative might be to allow young children to play with tangible physical devices, as children naturally do during the continuous development of their skills (Wang, Zhang, & Wang, 2011). Automated data collection then can be implemented e.g. by observing the participating child through computer vision. We, for instance, developed a setup where participants can play a game of Ludo using a physical game board. The game state was automatically captured through computer vision (Schmitt, Christopher, Tumanov, Weiss, & Möckel, 2018). Digitized game moves could be assessed and processed automatically in a computer, but the participants would not be affected by the technology. However, reliable digitalization of game states through computer vision is susceptible to lighting conditions and occlusions, requiring optimal conditions and making it difficult to use in regular classroom situations.

In this paper we thus explore the technology for yet another approach: we hide the required technology for digitalization and automated testing inside a physical game device. Inspired by the Porteus Maze test, a digital wooden maze device was developed that resembles those standard maze toys being used by children aged between 4 and 7. This digital wooden tabletop maze device is shown in Figure 4.1a. Inside the maze device, touch screen technology is hidden. The child moves a metal ball inside the wooden maze with a magnetic pen as shown in Figure 4.1c without realizing that the pen and ball movements are automatically recorded, making it possible to do these tests in a less formal, child-oriented setting. In this paper, we compare and analyze recordings automatically obtained by the wooden maze with recordings obtained from computer vision. For this, an overhead camera was installed as shown in Figure 4.1b, which tracked movements of the pink dot in the pen as visualised in figure 4.1d.

From tests with children, we show that the wooden maze can combine the advantages of both computer-based testing and classical standardized testing with physical objects. The digital wooden maze features automated data recording in real-time with high temporal and spatial resolution. Automated data analysis is performed offline on a standard PC. The maze is

a physical object that fits on a table and is easy to use by young children who treat it like a standard toy without facing technological boundaries.

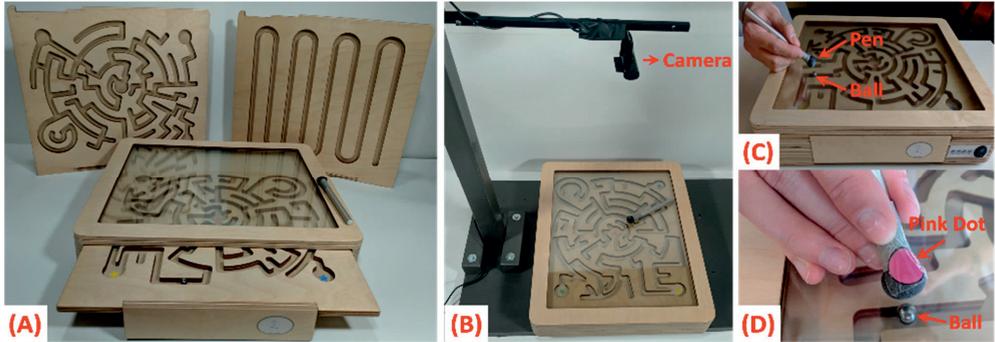


Figure 4.1 a. Digital Wooden Maze device with the different interchangeable maze structures, b. Digital Maze setup with overhead camera for computer vision experiments, c. Digital maze with custom-made pen, d. Custom-made pen for controlling the magnetic ball.

The remaining sections of this paper are organized as follows. Section 4.2 describes the digital wooden maze and data recording. Section 4.3 discusses the alternative approach of digitizing maze usage through computer vision. Section 4.4 and section 4.5 present and discuss experimental results, respectively.

4.2 Digital Wooden Maze: Hardware and Data Processing

The digital wooden tabletop maze device (shown in Figure 4.1a) weighs about 6.1kg. Its material costs less than 600€. In low-volume fabrication, a total price of less than 2500€ can be achieved. The maze device contains two separate touch screens placed back-to-back on top of each other. Screens are integrated into a wooden case as shown in the side view in Figure 4.2e. The wooden case and maze elements have been produced from sheets of Plywood that have been shaped through laser cutting and glued together. Inside the case and beneath the screens, the device contains a chamber that hosts the actual wooden maze structure. To allow for a variety of test items, we made the maze structures interchangeable. Figure 4.2b-d shows three existing wooden maze structures. The motor control structure (Figure 4.2d) contains a single trace and was designed to gain insight into the baseline (fine) motor capabilities of children. The assignment maze structure (Figure 4.2b) allows testing children with three different maze solving tasks of increasing complexity. The play maze structure (shown in

Figure 4.2c) contains a free play maze pattern with which we want to gain insights into unstructured play behaviour of children. The maze device contains four contact switches (Figure 4.2a) that allow for automatic detection of the inserted maze structure due to different patterns milled into the top left corners of each maze structure (marked in red in Figure 4.2b-d).

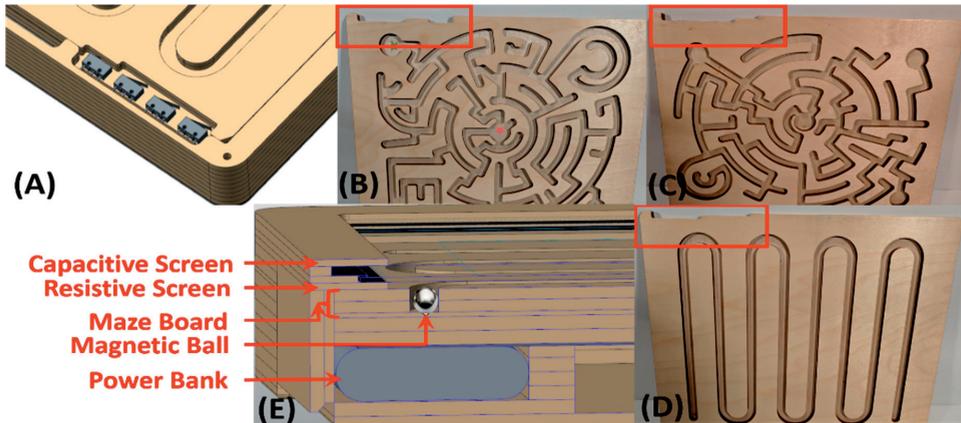


Figure 4.2 a. Contact switches inside the digital wooden maze device allow the automatic detection of inserted maze structures, b. a maze structure with different assignments, c. a maze structure for free play, d. a maze structure for motor control tasks. e. CAD picture of a cross-section of the digital wooden maze device.

A magnetic ball is inserted into the maze structures when sliding the maze structures into the digital wooden maze device. Once the maze structure is inserted, this magnetic ball can move freely inside the maze structure underneath the bottom screen but cannot accidentally be removed from the device.

Children move the magnetic ball within the maze structure with the help of a custom-made conductive pen (shown in Figure 4.1c-d). The pen was made from an anodized aluminium pipe of 12 cm height with conductive tip and designed so that it activates the capacitive multi-touch screen (BI-TFT19-PCAP10-4-USB from A1touch with an active area of 376.32 mm x 301.06 mm and a touch resolution of 4096x4096 pixels) on top of the maze device. As a result, when touching the top screen, the pen position can be continuously recorded by the maze device. In addition, the pen contains a magnetic tip that when close enough to the magnetic ball, pulls the magnetic ball against the resistive screen (5WR1902FA5 from A1touch with an active area of 376 mm x 301 mm and a touch resolution of 4096 x 4096 pixels) beneath the capacitive screen. As a result, the resistive screen can record the ball position. A

combination of a capacitive and a resistive screen was chosen, so that pen and ball position can be detected independently without affecting each other.

When calibrated correctly, by fusing the knowledge about the layout of the inserted maze structure with the pen and ball positions recorded by the capacitive and resistive screen, respectively, the position of the pen and ball with respect to each other and within the maze can be determined. In addition, it can be detected when the ball and pen are connected to each other and when a child moves the pen without the ball.

Figure 4.3 provides an overview of the main electronics of the digital wooden maze device: Data obtained from the capacitive and resistive touch screen are recorded and processed by an embedded STM32F429 processor board. For ease of use, this processor board can be controlled from a custom-made app via a Bluetooth wireless communication interface. To save data storage space and communication bandwidth, the processor board reads in the coordinate values of both the pen and the ball and stores these in a text file along with the time stamp following an event-based protocol. A new coordinate is registered if any of the following conditions is met:

- The X or Y coordinate of the pen changes by more than or equal to 2 mm.
- The state of the connection between pen and screen changes e.g. because the pen was disconnected from the screen.
- The state of the connection between pen and ball changes e.g. because the ball lost contact with the pen.

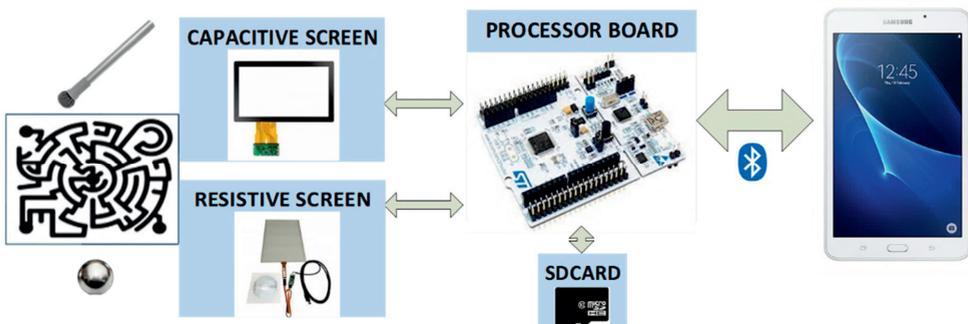


Figure 4.3 Overview of main electronics inside the digital wooden maze.

Recorded data are stored in the following format: time since the startup of the maze in ms; id of the registered touch [0..10]; X coordinate of pen in mm; Y coordinate of pen in mm; a binary value that indicates if the pen touches the screen [0,1]; a binary value that indicates if ball and pen are connected [0,1]; the id of the inserted maze structure [0..3]; X coordinate of ball in mm; Y coordinate of ball in mm.

To avoid any loss of data e.g. during a loss of communication, all recorded data are stored on an SD memory card inside the maze device. All electronics are powered by a 10400 mAh power bank allowing for more than 7 hours of continuous autonomous operation. The power bank is interchangeable so that the maze device can be kept operational during the charging process.

After recording, the data are transferred to a PC for automated post-processing, feature extraction, and visualization. Figure 4.4 provides an overview of the processing steps. After the data acquisition step, we remove all data points that are not of interest for us. Such recorded points e.g. can be caused by the touch screen detecting contact by the hand or elbow of a child. Filtering of points consists of 3 steps:

- 1) First, we remove all data points for which the ball and pen are not connected since for now we are only interested in movements where pen and ball move together.
- 2) Secondly, we remove those data points where the distance between the consecutive x and y coordinates is above a given threshold. This step allows for removing data points caused by other touch events than the pen. The ideal threshold was found heuristically through extensive experimentation.
- 3) Finally, we assign points according to the given tasks based on pre-defined start and end position on the maze as explained to the participating child.

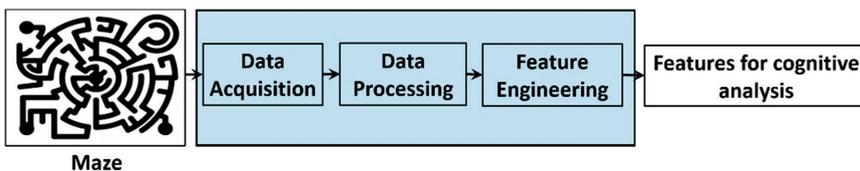


Figure 4.4 Overview of data processing stages.

From the filtered valid points, we can extract a variety of features. This includes both those features required by the Porteus maze test (like crossing maze lines or going into dead

ends) but also features that would be difficult to extract with a pencil-and-paper test like the position and duration that a child used to plan and rethink the applied maze-solving strategy. However, the evaluation and effect of these additional features for assessing cognitive abilities is beyond the scope of this paper. Here we focus on the presentation and evaluation of the new maze device mechatronics.

4.3 Digitalization of Pen Positions through Computer Vision

We track the position of the pen with a camera to compare the processed outputs of the digital wooden maze device against data from camera tracking. The experimental setup for the tracking process is shown in Figure 4.1b. Tracking was accomplished through an HD720p webcam running at 30fps and an open-source object detection algorithm in OpenCV ¹. For ease of tracking the pen, a pink dot was added to the tip of the custom-made pen (see Figure 4.1d) that is tracked during the experiments. Camera images are converted into HSV colour space for better detection performance. The tracking algorithm detects the presence of the coloured dot in the HSV camera images and saves the position to a text file. During the tracking process, the algorithm: (1) pre-processes the image frames (this step involves downsizing the image frame size and blurring the image to reduce high-frequency noise.), (2) localizes all the regions in the image frame within the specified HSV range, (3) searches for the radii size of the pink dot in the localized regions found in the previous step, and (4) saves the coordinates along with a timestamp to a text file.

We adjusted the colour threshold parameters by hand to achieve a good detection performance. Still, during experiments as expected from past studies (Schmitt et al., 2018) we experienced that detection performance was highly dependent on lighting conditions. To remove invalid point coordinates, we removed those data points where the distance between the consecutive x and y coordinates were above a given distance threshold. This step allows for removing falsely detected pen coordinates, similarly as for the data points directly obtained from the digital wooden maze device. Tracking ball positions was not possible with the camera setup due to occlusions and light reflections on the screen.

¹ <https://www.pyimagesearch.com/2015/09/14/ball-tracking-with-opencv/>, accessed 19 Mar 2019.

4.4 Experiments and Results

For an initial test of data recording quality, to compare data obtained from the screens and computer vision approach, and to generally demonstrate the feasibility of running tests with the digital wooden maze device, we invited three children of age 5 and one child of age 6 to test with us the device in two configurations. Children were asked to freely explore the maze equipped with the play maze structure and to run three assignments based on the coloured stars printed on the assignment maze structure (see Figure 4.5): Children always started at the centre of the assignments maze structure and then got asked (1) to go to the yellow star and back, (2) to go to the green star and back, and (3) to go to the blue star and back.

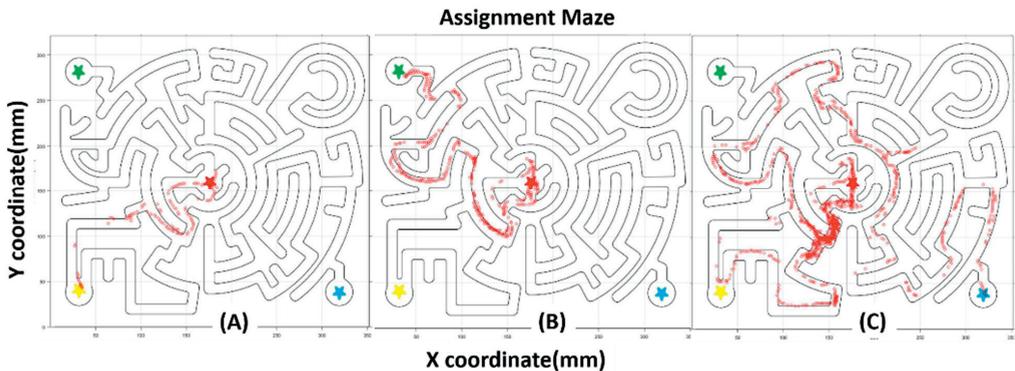


Figure 4.5 Drawing of assignment maze structure overlaid with extracted data points for the three different assignments. a. Assignment 1 (Go from red star into the middle to yellow star at the bottom left corner and back), b. Assignment 2 (Go from red star to green star in the top left corner and back), c. Assignment 3 (Go from red star to blue star in the bottom right corner and back).

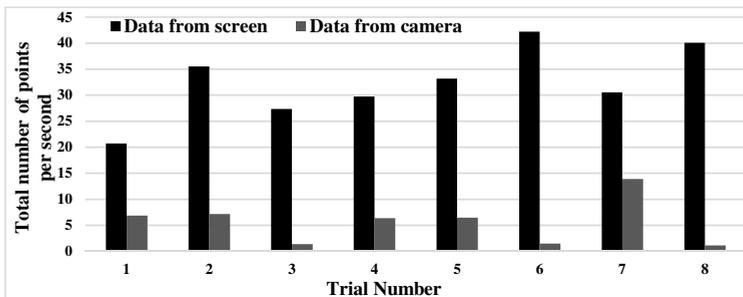


Figure 4.6 Temporal resolution of screen and camera approach: total number of recorded points per second.

Table 4.1 presents the total number of collected and filtered (invalid) data points for the four individual participants and maze structure experiments. Numbers are given for both the data obtained through the maze's screen and through computer vision (camera). For the computer vision experiments, we tried to create the best possible light conditions.

Table 4.1 Total number of data points collected and the invalid points from both the Maze screen and the camera.

Maze Structure	Trial No.	Participant No.	Maze screen			Camera		
			Total No. of points	No. of invalid points	Time (s)	Total No. of points	No. of invalid points	Time (s)
Assignment	1	1	9609	39	464.17	3232	27	473.68
	2	2	4589	52	129.24	852	32	119.60
	3	3	3326	37	121.65	145	25	105.33
	4	4	3349	44	112.62	769	334	120.84
Play	5	1	7419	18	223.67	1365	19	211.98
	6	2	2062	17	48.89	85	25	59.06
	7	3	3930	24	128.83	1620	63	116.59
	8	4	5935	48	148.26	169	80	149.50

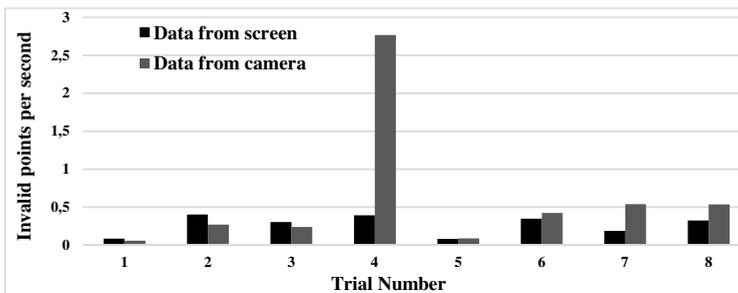


Figure 4.7 Temporal resolution of screen and camera approach: number of invalid points per second.

To compare the performance of our approach (screen approach) with the camera approach, we plot for each trial and approach: in Figure 4.6, the total number of points divided by the trial duration in seconds, in Figure 4.7, the number of invalid points divided by the duration of the respective trial in seconds, and in Figure 4.8, the number of invalid points divided by the total number of points in percentages. The screen approach provides on average 82.79% more points per second, in the best case 96.58%, in the worst case 54.45% and thus overall allows for a better temporal resolution. A reason is that the camera approach in contrast to the screen approach suffers from occlusions. So overall less points are detected in the camera images. In the screen method, inaccuracies are mostly generated when the pen loses contact with the ball. These usually occur if the pen is not held properly. As shown in Figure 4.7, in three out of eight trials, the screen approach produced a higher number of invalid points per second, on average 33.88%, and in the worst case 38.46%. This is partially caused by our definition of invalid points where we also filter situations where the ball is disconnected from the pen as invalid. This can also be caused by the fact that children sometimes rest their hands on the screen while playing, causing the top screen to detect both the hand and the pen. In the camera approach, the ball cannot even be detected. Thus, the data points that were classified as valid in the screen approach can be considered more reliable. The better screen resolution in

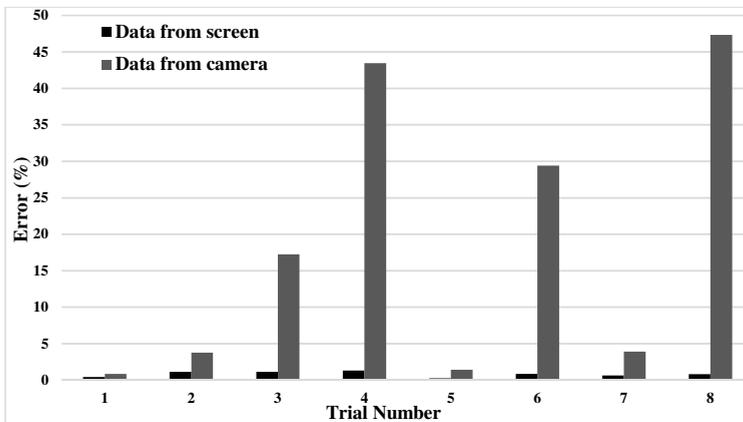


Figure 4.8 Screen resolution of screen and camera approach: Percentage of invalid (filtered) points with regard to the total number of points.

terms of valid detected points also becomes visible when calculating an error value by normalizing the number of invalid points with regard to the total number of points as displayed in Figure 4.8. On average the camera approach produces 95.62% more error than the screen approach, in the worst case 98.29% and in the best case 51.42%.

4.5 Discussion and Future Work

This paper presents a proof of concept for a digital wooden tabletop maze device designed for assessing the cognitive abilities of young children. The novelty of the presented material lies in the approach and hardware. The maze device allows for stealth assessments inheriting advantages of computer-based testing like automated data acquisition and analysis but with a key advantage: The maze device looks like a normal wooden maze. Children can interact with the maze device in the way they are used to do with other tangible physical devices like toys. To proof the value of the maze device as a cognitive assessment tool, we are currently conducting a representative study with more children. So far, this paper presents only a pre-study with four children. However, the fact that the maze device allows for an extraction of similar features as the well-known Porteus maze test and mazes subtest from the WPPSI-R test is promising. Also, initial feedback from the four children who tested the device is very promising: all children showed great excitement when playing with the device and had no problems in handling the device. Also, their parents showed strong interest, making us hopeful that the maze device will be well-accepted by a general population.

Inspired by our experience with our camera-based digitalization approach of a board game for cognitive assessments, we compared the results obtained from the screens that have been integrated into the maze device with a camera-based tracking approach. Even under well-adjusted lighting conditions, the screen approach outperforms the camera-based approach. Due to its screen technology combining capacitive and resistance sensing, the digital maze is not susceptible to lighting conditions at all but can operate robustly in less-structured environments.

4.6 References

- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x
- Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning disabilities, 43*(6), 541-552.
- Borghans, L., Meijers, H., & Ter Weel, B. (2008). The role of noncognitive skills in explaining cognitive test scores. *Economic inquiry, 46*(1), 2-12.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616.
- Fleege, P. O., Charlesworth, R., Burts, D. C., & Hart, C. H. (1992). Stress begins in kindergarten: A look at behavior during standardized testing. *Journal of Research in Childhood Education, 7*(1), 20-26.
- Kaufman, A. S. (1990). The WPPSI-R: You can't judge a test by its colors. *Journal of School Psychology, 28*(4), 387-394.
- Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. New York: John Wiley & Sons.
- Kirsch, P., Lis, S., Esslinger, C., Gruppe, H., Danos, P., Broll, J., . . . Gallhofer, B. (2006). Brain activation during mental maze solving. *Neuropsychobiology, 54*(1), 51-58.
- Pillay, H. (2002). An investigation of cognitive processes engaged in by recreational computer game players: Implications for skills of the future. *Journal of research on technology in education, 34*(3), 336-350.
- Porteus, S. D. (1959). *The Maze Test and clinical psychology*. Oxford, England: Pacific Books.
- Schmitt F., Christopher S.M., Tumanov K., Weiss G., Möckel R. (2018) Evaluating the Adoption of the Physical Board Game Ludo for Automated Assessments of Cognitive Abilities. In: Göbel S., et al. (Eds.), *Serious Games. JCSG 2018. Lecture Notes in Computer Science, vol 11243* (pp.30-42). Cham: Springer International Publishing.

Wang, D., Zhang, C., & Wang, H. (2011). T-Maze: a tangible programming tool for children. *Proceedings of the 10th International Conference on Interaction Design and Children*, pp. 127-135. doi:10.1145/1999030.1999045



Chapter 5

Children's executive functioning: What can solving mazes using Clever Maze reveal?

This chapter is based on joint work with Seethu Christopher^a, Karien Coppens^b,
Petra Hurks^c, Rico Möckel^a and Lex Borghans^b.

^a Department of Data Science and Knowledge Engineering, Faculty of Science
and Engineering, Maastricht University

^b Department of Macro, International and Labour Economics, School of Business
and Economics, Maastricht University

^c Department of Neuropsychology and Psychopharmacology, Faculty of
Psychology and Neuroscience, Maastricht University

5.1 Introduction

Observing children's behaviour during task (or test) performance provides information on their cognitive development. Collecting this information via tests and systematic observations can help to detect potential cognitive strengths and/or weaknesses and allows for tracking of cognitive skill development. Particularly in early childhood, when functional and structural brain development is still taking place (Brown & Jernigan, 2012), these insights are highly valuable, which emphasizes the importance of testing. In early childhood, such as kindergarteners, testing, or assessment, is difficult to perform, however. For one, testing requires standardized, potentially strenuous situations for children, and these children are asked to perform tasks that are often not aligned with their interests.

In this study, we explored if we could capture cognitive functions, and more specifically executive functions (EF, for definition, see below) of kindergarteners, by letting the children make assignments with a toy. For this purpose, we believed technological means could be applied to provide an innovative approach that is more suitable for testing cognition in (young) children. We have developed a toy with embedded sensors for kindergarteners, namely Clever Maze (CM), which is described in Chapter 4. In this chapter we related data collected with this tool to the performance of children on traditional, pencil-and-paper, EF tests to investigate whether a valid understanding of EF can be obtained with CM. First, we will define more clearly what EF are and why they are important. Next, we will discuss the traditional EF tests and the role of technology. Third, we explain our study and elaborate on how technology and sensors can, potentially, be used to enhance toys for measurement purposes.

5.1.1 The Differentiation of EF in Children

EF is an umbrella term for describing multiple, interrelated functions needed for goal-directed behaviour (Best & Miller, 2010). Understanding EF in children is highly valuable, as individual variations in EF can be meaningful and predictive of a variety of important life outcomes. A relation between EF and school readiness has for instance been found in preschoolers (Blair & Razza, 2007; Morrison, Ponitz, & McClelland, 2010), as well as school success in e.g., 10-11 year olds (Borella, Carretti, & Pelegrina, 2010) and physical health in e.g., teenagers (Miller, Barnes, & Beaver, 2011).

There is consensus that, in adults, EF consists of three basic cognitive components, namely working memory, inhibitory control, and cognitive shifting (Diamond, 2013; Lehto,

Juujärvi, Kooistra, & Pulkkinen, 2003; Miyake et al., 2000). Diamond (2013) summarized that inhibitory control refers to the ability of the individual to overcome internal urges or responses to external lures via controlling one's attention, behaviour and/or emotions, so one is able to respond appropriately to external stimuli. Working memory is defined as keeping information active in the mind to be able to work with it. Lastly, cognitive shifting or cognitive flexibility refers to being able to shift perspectives, as well as being able to adjust to unexpected demands or situations. These basic EF processes allow for higher-order EF processes, such as reasoning, problem-solving and planning (Collins & Koechlin, 2012; Lunt et al., 2012).

In adults, the three basic cognitive components (i.e. inhibition, working memory and shifting) have been found to be interrelated yet separable, via means of structural equation modelling (Miyake et al., 2000). However, during childhood, when cognitive development is ongoing, the differentiation between components is different than in adults and is subject to developmental changes. Indeed, Viterbori, Usai, Traverso, and De Franchis (2015) conducted for instance a confirmatory factor analysis studying the latent EF-structure in kindergarteners (i.e., children aged 5-6 years), and they found that a two-factor model, consisting of an inhibition factor and a working memory/shifting factor, best fitted their data. From this, we can conclude that when studying the EF of kindergarteners, we can expect to be able to measure working memory and inhibition as unique basic cognitive components, but not cognitive shifting. This should be kept in mind when studying the relationship between CM and the outcomes on the traditional, paper and pencil EF tasks.

5.1.2 EF Tests and the Role of Technology

Traditionally, paper-and-pencil EF tests for children have been derived from neurocognitive tasks or tests for patients, and often adult patients, with frontal lobe damage (Carlson, 2005). Even in these populations, Chaytor and Schmitter-Edgecombe (2003) critically stated that the ecological validity of these EF tests, i.e. the degree to which test performance is representative for EF performance in everyday life, was often only moderate. One reason for this may be related to the manner in which test performance is scored. For example, in multiple, traditional EF tasks, multiple items are administered. An answer to an item is scored as correct if the correct response is provided within a certain time limit and testing is often stopped after a consecutive number of wrong answers. Test performance is then calculated by summing the total number of correct answers. This obtained end score (i.e., the total number of correct items)

is then used for (clinical) interpretations, leaving out potentially valuable information on the processes underlying this obtained end score. For example, variations within the time needed to solve the assignment could potentially provide valuable insights into cognitive functions in general, such as processing speed, and EF, such as inhibition, more specifically.

Designing a toy that can collect data automatically can yield several advantages. Since sensors can collect detailed and continuously collected (over time) information on cognitive processes that are usually discarded in traditional EF tests, detailed insights could be achieved. As these processes could be ultimately automatized, less effort from the examiner would be required i.e. less time and energy would have to be invested in testing and scoring, making wider application possible. For instance, the potential increase in ease of assessment implies that data can be collected in larger samples and can be done in varying contexts other than the one-to-one setting. With sufficient testing and research, a less standardized situation could be created as well, as children could be encouraged to try and solve mazes with CM. All of these benefits combined have the advantage of potentially leading to a less strenuous testing situation and more informative data on the cognitive development of children. Receiving such detailed and continuously collected information on task performance is, for instance, difficult when scoring in a pencil-and-paper format.

In a way, tasks performed on computers or tablets already provide an opportunity for tackling this issue. Several tests have been developed in this format, using simple or elaborate game-like designs to appeal to children. For example, Tsuchiya, Oki, Yahara, and Fujieda (2005) used a computerized version of the Wisconsin Card Sorting Task for children in their study instead of the traditional version that uses flashcards. Though the comparison to the traditional version was not explicitly made in their discussion, they specified their choice for a computerized version in order to limit the verbal cognitive function, as well as social and verbal demands because part of their sample focused on children with autism. Another example is given by Józsa, Caplovitz-Barrett, and Morgan (2017) who developed, amongst other things, a tablet application to measure EF in 3- to 8-year-olds, with reliable and valid measures. They argued that by using an app, data collection would require less teacher training or time, though they did not make an explicit comparison to pencil-and-paper versions of similar tasks with respect to task performance and validity is to be tested in follow-up studies.

Besides computers and tablets which are more in line with our study, toys that are technologically enhanced have also been used for the purpose of EF stimulation. Di Lieto et al. (2017) example studied the use of educational robots as a means to promote EF. The authors studied 5- to 6-year-olds, focusing on these children's EF (working memory and inhibition) as outcomes, as well as their visuospatial abilities and attention. The results revealed that a short, intensive training of programming a small robot (the Bee-Bot) led to an improvement (post-training) in working memory and inhibition skills, but not in attention and visuospatial abilities. The usage of real-life tools, instead of screen-based methods, has the advantage that tools actually exist in the space and can be manipulated manually by the children. We believe this is a valuable aspect and a unique benefit of sensor embedded toys.

5.1.3 The Current Study: Sensors in a Maze Toy

We developed a wooden maze toy called CM, which includes sensors to track information on how children solve maze patterns (see Chapter 4). The child controlled a ball under a glass-looking surface with a magnetic pen. CM had an assignment condition in which children had to get the ball to different locations in the maze in several assignments. The target group of the study were kindergarten children who were on average 6 years old, ranging from 64 to 86 months old. The data collection by CM allowed objective measurement of the task behaviour of children, while they were solving a maze. We extracted variables related to the process of solving the assignments, such as the number of mistakes. We then tested the validity of the obtained CM data. More specifically, we studied the convergent validity (i.e., the degree to which scores of two tests, that theoretically should measure the same construct(s), are significantly correlated) and the discriminant validity (i.e., whether scores of two tests that theoretically measure different constructs are indeed unrelated). Significant and insignificant correlations between EF tasks and CM variables will be discussed in this framework. Our hypothesis was that a range of meaningful variables can be obtained from the data collected by the digital maze, in the sense that these variables contain sufficient variation between children and are a reflection of underlying EF. We expected some relation with tasks that require children's inhibition and working memory skills, but not yet cognitive shifting. In the study of Di Lieto et al. (2017) a similar target group was studied and a broader EF test battery was included (i.e. not only inhibition and working memory). We adopted this approach as well, since solving maze tasks requires various EF for which we included a broader cognitive test battery. Furthermore, we also added tests we expect to find weak relations to the performance

on CM in order to establish discriminant validity, such as tasks with a stronger language component.

5.2 Method

5.2.1 Participants

Elementary schools in the Southern region of the Netherlands were approached to participate in the study. If they agreed to participate in the study, children of these school in the second year of kindergarten were invited to partake via a letter to their parents. In this grade, children were on average 6 years old and normally start formal primary school the following school year (i.e. commonly known as Grade 1). Parents received an information letter and informed consent form. Children were invited to participate if their parents gave permission. Forms were only distributed to children who were eligible for the study, which was decided based on teacher judgment on several predefined criteria. More specifically, we excluded children who had repeated a grade, but also children who did not have at least one parent who sufficiently understood Dutch, since all information was provided in this language. Finally, we also did not include children with a diagnosed disorder such as a developmental or psychiatric disorder. The criteria were included in order to focus on the population of children within a normal range of development, rather than having developmental or language issues as a potential confounder in the analyses.

5.2.2 Procedure

Materials.

Children were tested individually in an empty, relatively quiet room of their school. Two testing sessions were planned for each child in order to avoid making testing strenuous, each session lasted on average 30 minutes. The testing occurred in a one-to-one setting. During the first session CM, the Raven CPM, Five-Point test and Mazes were administered. During the second session Tower of London, Verbal Fluency test and Digit Span were conducted.

Clever Maze (CM).

Description of CM. We designed CM for children to play with, in which they can move a magnetic ball under glass with a magnetic pen. The toy had no outside wires and looked like a regular wooden toy (see figure 5.1). On the back, boards could be switched and LED-lights provide functional feedback on the back of CM, which was only visible to the researcher (see figure 5.2). CM was however capable of collecting data on the play of the children, in particular time and the coordinates of the ball. Time and 2D-coordinates of the pen were collected with an indicator of whether the ball is attached to the pen, allowing tracking of the movement of children through the maze. The magnetic ball could not exit the digital maze and could only be manipulated by moving the magnetic pen over the glass surface. Children could not manipulate the orientation or tilt of the digital maze itself and were encouraged to stay in the same position facing the maze.

The technical specifications of CM are described in-depth in Chapter 4.

For this study, we used data from an assignment board which was inserted in the maze. There were four points of interest indicated on the board with a coloured star (one in the centre and three in outer corners). Each assignment consisted of getting the ball from the central star to another star in a corner (part one) and back (part two) (see figure 5.3 and 5.4). The three assignments with each two subparts made 6 assignments. Difficulty level was determined by the distance needed to solve the task and number of dead ends children encounter before arriving at the end destination of the assignment. For the first assignment, there were two



Figure 5.1 Picture from the researcher's perspective of the backside of CM, showing how the tool is controlled and board can be switched.



Figure 5.2 Picture from the child's perspective of CM with pen.

possible and correct roads to get to the answer and only two potential dead-ends the children should avoid. For the second assignment, the distance to solve the maze was longer and aside from the part that overlaps the first assignment, there was only one route to get to the answer. The child had to bypass six dead ends to get to the answer. For the last assignment, there were 21 potential dead-ends the child should avoid in order to complete the assignment. There were two instances where a split in the route did not lead to a dead-end but still allowed the child to stay on track, which can be seen as a slight sidetrack to the fastest route. Other than that, there was only one route to the star in the corner.



Figure 5.3 Pictures from the researcher's perspective displaying the three assignments, with the first assignment on the left, medium difficulty in the middle and most difficult on the right.

Children could focus on the assignment at hand since parts of the maze were covered up and later unfolded as assignments became more difficult, as can be seen in figure 5.3. For the first assignment, only the top right quadrant from the perspective of the child was visible. The child was asked to move the ball from the central star to the one in the corner and back. The second assignment revealed the bottom right quadrant of the full maze by removing the smaller cover panel and the assignment was to get the ball from the central star to the new corner star and back. The third assignment required the whole maze, where the child had to get the ball from the central star to the new corner star and back.

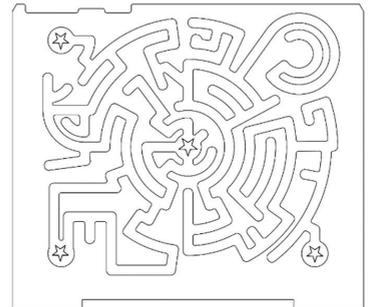


Figure 5.4 Graphical depiction of the maze design, with the position of the four stars indicated.

CM testing procedure.

Protocol. A detailed testing protocol was devised to ensure standardized testing of the children with the digital maze, which can be found in Appendix C. It included a description of how to start up and shut down CM and control it with the external device, as well as exact

verbal and visual (e.g. pointing to the star where the child should go) instructions that should be given to the child. All children went through the protocol in the same fashion. Before the child entered the room to start the testing procedure, the maze was set up in a certain way, so the pattern of the maze was not yet visible upon entrance. CM was placed on a children's table that allowed the children to stand while they would play with it, thus ensuring sufficient visual overview of CM. The child was to maintain the same orientation toward CM. The researcher would sit across from the child.

For each assignment, the researcher explained that the ball is placed at a certain star and that the child should get the ball to another star, by using the pen. The researcher named this star by colour and points to it on the maze design where the child is supposed to go. Immediately after the provision of this short instruction, the start button was pushed to start the measurement, as well as a marker indicating in the data that the following data contain a specific assignment. From the moment the child started moving the ball with the pen, data were continuously collected and stored. When the child completed the first part of an assignment, the second part was to go back to the starting point. A second marker was pushed as the researchers verbally and visually explained that the child should go back now to the initial star. As assignments became more difficult and reveal more of the maze, the wooden cover panels were removed until every assignment was completed.

Criteria for discontinuation. All children started with the first assignment. They were allowed to make the second and third assignment if the first assignment was completed successfully. Only if the child could not successfully complete the first and second assignment on two attempts (i.e. they could not reach the endpoint successfully), the child did not have to make the third assignment. If children froze or did not know how to continue during an assignment, the researcher encouraged them and gave them one minute (assignment 2) or 2 minutes (assignment 3) time to think before discontinuing the assignment. Children were encouraged gently if they indicated they did not know the answer or did not want to continue, but if no progress was made afterwards, the assignment was discontinued. All encouragements were superficial and unrelated to the content of the task, e.g. 'Keep trying'.

Data extraction of assignment maze.

Measurement data processing was carried out in a systematic order using Matlab 2017b. Figure 5.5. shows the data flow in steps.

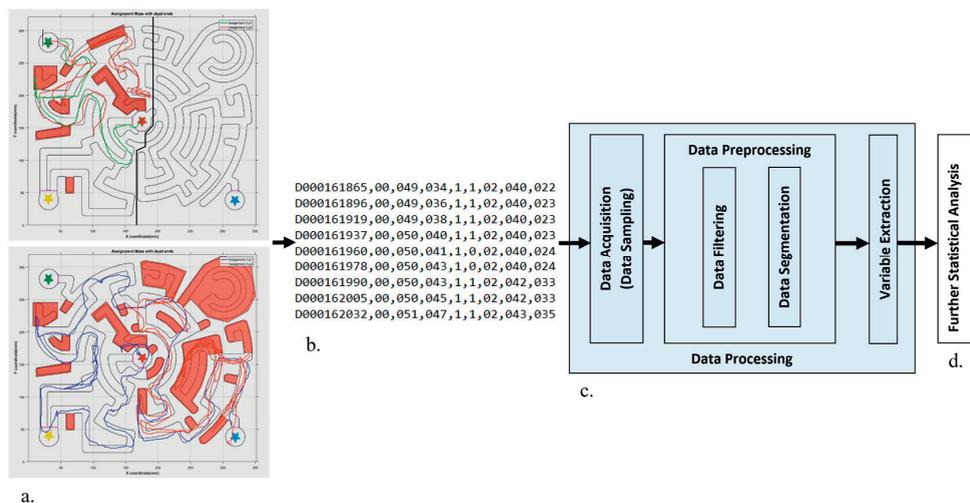


Figure 5.5 Data flow measurement data processing. a. Depiction of the movement of the child through CM. The red shaded areas indicate dead ends in the maze, b. Illustration of the raw data provided by CM, c. Schematic depiction of the data processing steps, d. The resulting variables are used for further statistical analysis.

We utilized the raw data of CM to create process measures. First, the distance children needed to solve the assignment, expressed in centimetres. If children took detours or entered dead ends, a longer distance was needed to solve the assignment. Second, the number of dead ends they enter while solving the (sub-) assignments was accounted for. Third, we extracted inactive time in seconds, which was defined as time children stood still in the maze, for example, to think or because they didn't know how to proceed. Fourth, their average speed of moving was collected for which only moving time was included (not time standing still).

Paper Mazes (PM). Participants performed the Mazes task, which is a subtest of the WPPSI-R (Vander Steene & Bos, 1997). Children had to solve several mazes with increasing difficulty. They were scored on whether they solve the maze within a certain period without making too many errors, such as entering blind alleys or crossing walls. Only within the allowed timeframe and within a predefined maximum number of errors allowed per assignment, a maze would be scored as 'pass' (score: 1) rather than 'fail' (score: 0). The score assigned to a passed maze solution (i.e. if the child solved the maze) depended on the number of mistakes made under the maximum number, with the highest score assigned to mazes solved without any mistakes. The score (maximum 26) was transformed into a normed score (ranging from 1-19) according to the scoring manual of the test, which was used in the analysis. The subtask Mazes has a split-half reliability of between .77 and .85. Stability of testing, i.e. test-

retest reliability, was below .60 (Kaufman, 1990). Lezak et al. (2004) mention that this task is often used as a practical and satisfactory substitute for the lengthier Porteus maze test (Porteus, 1959). The authors summarize from multiple findings that solving maze tasks is often related to planning and visuospatial abilities in adults (Ardila, Rosselli, & Rosas, 1989; Daigneault, Braun, & Whitaker, 1992). However, we also established in the introduction that planning abilities are more difficult to measure in young children. In order not to confuse this task with outcomes from CM, we refer to this task as Paper Mazes or PM.

Raven Coloured Progressive Matrices (RCPM). The RCPM can be assessed in children aged 5 to 11 years (Raven et al., 1998) and does not require a verbal response from the participants. Lezak et al. (2004) states that the RCPM measures visuospatial reasoning and problem-solving. Children were shown a picture of a pattern with a piece missing from it, with at the bottom of the page six options for filling in the missing piece. Children had to choose the piece they believed to be completing the pattern. Each potential piece had a number. Children could state their response verbally to the researcher or point to their answer. The test consisted of three sets within a test booklet, A, Ab and B, which increase in difficulty. Each child went through all sets. Only when the child made mistakes in the first five items of each set, the researcher could refer back to the first item made and explain the aim of the task again. Afterwards, all responses were noted, regardless of their correctness. The score was the total number of correct answers. Raven et al. (1998) reported originally low test-retest reliabilities of .65 for children under the age of 7, but also that newer studies yield satisfactory reliabilities when assessed by split-half or alternative test-retest methods. The minimum score was 0, the maximum score was 36.

Tower of London (ToL). The Tower of London task stems from the NEPSY test battery (Korkman et al., 1998) originally designed by Shallice (1982). This test is specifically designed for children from 5 to 12 years old. Tower tasks, similar to the ToL, are believed to measure e.g., planning and problem solving (Unterrainer et al., 2004), as well as (visuospatial) working memory, response inhibition and visuospatial memory (Carlin et al., 2000; Phillips, 1999; Welsh et al., 2000) in adults. Split-half reliability ($r = .72$) and internal consistency ($\alpha = .69$) have been shown to be satisfactory in adults (Kaller, Unterrainer, & Stahl, 2012). The test consisted of moving three balls on three pegs in such a way it is a representation of a picture shown to a child. An answer was scored as correct if a child solved the puzzle with a specific number of steps within a specific time limit. Children were only allowed to move one ball at a

time and had to leave the rest on the pegs when moving a ball. 20 items were administered unless children could not complete four successive items correctly. Maximum raw score was 20, but for the study, we use the scaled score with a maximum of 19.

Five-Point test (FPT). This task serves as a figural fluency test, which is a nonverbal analogue to a verbal/word fluency test. The test was designed by Regard et al. (1982). Tucha et al. (2012) tested children and teenagers within the age range of 8 to 15 years. They reported strong relations between the number of unique designs and processing speed and mental flexibility, and medium relations with short-term memory, problem-solving and inhibition. However, the children in our sample are significantly younger and flexibility is not yet a separated EF in kindergarteners, as discussed in the introduction. Children received a sheet of paper containing 40 boxes, within each box five dots. Children had to draw as many unique designs as possible by connecting at least two dots with straight lines for which they received one minute time. No figures were to be repeated and only single lines were to be used, in line with the protocol of Strauss et al. (2006). Before the test, two sample drawings were made by the researcher to illustrate the principle. The child practised two drawings and received feedback and another chance to practice in case mistakes were made. Each drawing was scored as either correct, false (e.g. using curved lines) or double. Psychometric properties on validity and reliability have been tested and shown to be in order (Tucha, Aschenbrenner, Koerts & Lange, 2012). There was no pre-set maximum score, as children could even receive a second page if the first was full. For this study, we utilized the total number of unique designs as the main FPT outcome measure.

Verbal Fluency task (VFT). This task is an adaptation of the NEPSY version (Korkman et al., 1998). We utilized the animal naming version of the task, which can be seen as a categorical verbal fluency or semantic fluency task (Tombaugh, Kozak, & Rees, 1999). The task has been shown to correlate with verbal memory and language comprehension (Ardila, Ostrosky-Solís, & Bernal, 2006). The aim of the test is to search internal lexicon rapidly and automatically (Riva, Nichelli, & Devoti, 2000), suggesting a language component for solving the task (Whiteside et al., 2016). Test-retest reliability was established at .68 for the semantic fluency (animal naming) test in adults (Harrison, Buxton, Husain, & Wise, 2000). Children had to name as many animals as they know, as fast as they could within a one-minute time span. Two examples were named by the researcher, namely cat and dog. Responses were noted by the researcher and were later divided into number of correct animals, false animals

(e.g. fantasy animals such as dragons) and doubles. For this study, we utilized the total number of correctly named animals, which is what shall be referred to from now on. Minimum score is 0, there is no maximum score predetermined.

Digit Span forward and backwards (DSfw, DSbw). This task is a subtest of the WISC-III (Kort et al., 2002). The researcher read numbers aloud and the children had to repeat them. Difficulty increased per two items, by adding another number that was read aloud, starting with two numbers in the first round. Afterwards, children had to repeat number sequences backwards, which is the reverse condition. The total score was the number of correctly (backwards and forward) repeated number sequences. The maximum score for the forward condition was 16 and for the backward condition 14. The reliability coefficient of the Dutch digit span task was .59 for 6.5-year-old children. Split-half reliability was .79 for the same age group and test-retest reliability was .67 for 6 to 8-year-old children (Kort et al., 2002). Lezak et al. (2004) summarized that adding the scores of the two tasks assumes they measure the same underlying construct. The authors argued this is not true, at least for adults, as the forward variant is presumed to measure attention (Kaufman, McLean, & Reynolds, 1991), whereas the backward variant measures working memory (Banken, 1985). For this reason, we treated both tests separately in this study.

Go/No-go task: no-go accuracy (NGA). This task is widely used as a measure for inhibition (Johnstone et al., 2007; Jonkman, 2006) as well as attention and/or working memory processes (Criaud & Boulinguez, 2013). Langenecker, Zubieta, Young, Akil, and Nielson (2007) stated that the psychometric properties of the Go/No-go task were insufficiently established. We programmed a Go/No-go task in Presentation software version 20.1. Children could perform the task while looking at a laptop screen that was placed in front of them while providing responses to stimuli on an RB-844 response box (Cedrus Corporation). We showed children a picture of a cartoon dog and cat on two separate sheets of paper. We asked them to point to the dog and then to the cat to ensure they knew the difference. They then received 15 practice trials in which the go stimulus (cat) and no-go stimulus (dog) are shown in random order on the laptop. Stimulus duration was 250ms and inter-stimulus interval was 1500ms. Children received a green smiley face if they responded correctly (i.e. pressing the button when the cat is shown, not pressing the button when the dog is shown) and a red frowning face when they responded incorrectly (i.e. pressing the button when the dog is shown, not pressing the button when the cat is shown). The feedback face was shown for 2000ms. The practice trial

should have 40% correct responses in order to proceed to the actual task. The task contains 96 go stimuli and 48 no-go stimuli that are shown in random order. No feedback was provided. Stimulus duration was 250ms and the interstimulus interval was 1500ms. For this study, we utilized the accuracy with which children succeeded in inhibiting their responses when the no-go stimulus was shown.

5.2.3 Analyses

Descriptive statistics are provided on the sample, the EF test battery outcomes, and the CM variables. We performed correlational analyses using Pearson's *R*. Next, we performed factor analyses on the subgroups of CM variables. We extracted the factors to be retained and correlated them to the EF test battery again using Pearson's correlation.

5.2.4 Ethics

The Ethical Review Committee of Psychology and Neuroscience at Maastricht University (ERCPN-175_02_01_2017) approved this study.

5.3 Results

5.3.1 Descriptive Statistics

All analyses were performed in Stata version 15.1. In total, 121 consent forms were positively filled out by parents. In the analyses, 103 children were included. The 18 excluded children were either absent from school on testing days or there were problems in the data collection due to technical issues with the maze leading to missing data. For example, if the Bluetooth connection was lost, no data were collected. Outliers on all outcome variables of the EF test battery as well as the CM variables were replaced with a value of 3 standard deviations plus or minus the mean of that variable. This restriction was applied in all analyses. The sample included 51 boys and 52 girls. The children were on average 72.5 months old (6 years) with 4.4 months standard deviation. Because of practical reasons, the Go/No-Go task was added later in the data collection and is therefore available for a smaller group of the sample ($N = 51$). The number of children for each outcome fluctuated since only outcomes on successfully completed mazes were included in the analysis. Table 5.1 provides descriptive statistics on the CM variables. There are four categories of CM variables, namely distance, inactive time, average speed and number of dead ends entered. For each of the categories, six variables were

extracted, one per sub-assignment. Table 5.2 provides descriptive statistics on the EF test battery.

Table 5.1 Descriptive information on the variables collected by CM.

CM variables	N	M	SD	Min.	Max.
Distance 1.1	102	32.22	7.92	6.91	57.58
Distance 1.2	101	33.61	8.29	10.99	62.06
Distance 2.1	103	94.20	40.62	47.42	230.62
Distance 2.2	102	65.09	20.01	40.37	131.41
Distance 3.1	98	262.00	128.65	124.83	685.49
Distance 3.2	100	208.47	95.32	120.06	547.68
Inactive time 1.2	102	0.03	0.15	0	0.88
Inactive time 2.1	103	0.69	1.97	0	8.25
Inactive time 2.2	102	0.08	0.49	0	4.02
Inactive time 3.1	98	4.62	9.34	0	58.29
Inactive time 3.2	100	0.75	1.69	0	8.32
Avg. speed 1.1	102	11.15	2.95	4.41	18.94
Avg. speed 1.2	101	12.17	4.54	5.98	34.86
Avg. speed 2.1	103	12.05	2.86	6.96	22.35
Avg. speed 2.2	102	14.32	7.71	8.06	56.14
Avg. speed 3.1	98	14.84	7.43	7.56	52.83
Avg. speed 3.2	100	14.03	3.63	9.09	35.32
Dead ends count 1.2	102	0.02	0.09	0	0.54
Dead ends count 2.1	103	1.84	1.32	0	6.17
Dead ends count 2.2	102	0.83	0.94	0	3.82
Dead ends count 3.1	98	4.26	2.95	0	13.18
Dead ends count 3.2	99	4.05	3.05	0	15.25

Note. Distance variables are measured in centimetres, inactive time variables are measured in seconds. For the first part of the first assignment, no children exhibited thinking time or entered dead ends. These variables are therefore left out. 'Distance 1.1' refers to the distance travelled to solve the first part of the first assignment. Similar naming strategy is used for all CM variables.

Table 5.2 Descriptive information on the EF test battery.

EF test battery	M	SD	Min.	Max.
ToL (N=103)	13.16	2.41	5.24	18
DSfw (N=103)	6.17	1.78	3	11
DSbw (N=103)	2.67	1.14	0	5
FPT (N=103)	7.11	2.68	1	14
RCPM (N=103)	23.21	4.34	10.08	34
VFT (N=103)	10.58	3.67	3	21
PM (N=101)	10.48	2.52	6	18.12
NGA (N=51)	0.68	0.17	0.29	0.98

5.3.2 Correlational Analyses

Pearson's pairwise correlation was utilized to understand the relation between the CM variables and the EF test battery outcomes, see table 5.3 for results. In interpreting table 5.3, we looked at groups of significant correlations. Such groups could be indicative of some relation between the type of measure from CM and the EF outcomes that deserved further investigation. On the EF tasks, a higher value is indicative of better performance. With regard to CM, higher values for distance are indicative of inefficient route use or mistakes being made. The same rationale was used for the number of dead ends entered. Average speed will be higher if children move faster through the maze, whereas a longer inactive time can lead to more time needed to solve the mazes, or lead to higher efficiency if the inactive time is used to form a strategy (instead of trial-and-error maze solving). Table 5.3 indicates that the distance of CM had two significant and two slightly significant negative correlations with the ToL outcome. The RCPM had the same number of relations with CM distance travelled and three highly significant negative correlations with the number of dead ends entered. Performance on the PM had two highly significant negative correlations with the number of dead ends entered. DSfw had one significant and one slightly significant correlation with distance as well as the number of dead ends entered. The rest of the significant correlations were not grouped and thus not further discussed into detail.

Table 5.3 Correlation table displaying Pearson's pairwise correlations between EF test battery outcomes and CM variables.

CM variables	EF tasks							
	ToL (N)	RCPM (N)	FPT (N)	VFT (N)	DSfw (N)	DSbw (N)	NGA (N)	PM (N)
Distance 1.1	-.18* (102)	-.16 (102)	-.13 (102)	-.06 (102)	-.03 (102)	-.19* (102)	-.16 (51)	.09 (100)
Distance 1.2	-.23** (101)	-.09 (101)	-.03 (101)	.00 (101)	.08 (101)	.04 (101)	-.25* (50)	.05 (99)
Distance 2.1	-.10 (103)	-.23** (103)	-.12 (103)	-.01 (103)	-.19* (103)	-.12 (103)	.16 (51)	-.04 (101)
Distance 2.2	-.19* (102)	-.24** (102)	-.02 (102)	.24** (102)	-.22** (102)	-.18* (102)	-.16 (50)	-.13 (100)
Distance 3.1	-.24** (98)	-.20* (98)	-.12 (98)	-.12 (98)	-.06 (98)	-.17 (98)	.04 (50)	-.23** (96)
Distance 3.2	-.09 (100)	-.19* (100)	-.05 (100)	.06 (100)	.00 (100)	-.05 (100)	-.12 (50)	-.04 (98)
Inactive time 1.2	-.01 (102)	.03 (102)	-.03 (102)	-.06 (102)	.22** (102)	.16 (102)	. (51)	.13 (100)
Inactive time 2.1	.08 (103)	-.23** (103)	-.14 (103)	.14 (103)	-.09 (103)	-.16 (103)	.07 (51)	.04 (101)
Inactive time 2.2	-.18* (102)	-.25** (102)	-.16 (102)	.09 (102)	-.11 (102)	-.32*** (102)	.02 (50)	-.05 (100)
Inactive time 3.1	-.06 (98)	-.05 (98)	-.27** (98)	.13 (98)	.00 (98)	-.07 (98)	.18 (50)	-.06 (96)
Inactive time 3.2	-.04 (100)	-.21** (100)	-.16 (100)	.04 (100)	-.11 (100)	-.07 (100)	.10 (50)	-.04 (98)
Avg. speed 1.1	.12 (102)	.08 (102)	-.16 (102)	.03 (102)	-.02 (102)	.02 (102)	-.08 (51)	-.15 (100)
Avg. speed 1.2	.08 (101)	.15 (101)	.14 (101)	.15 (101)	.00 (101)	-.01 (101)	.06 (50)	.02 (99)
Avg. speed 2.1	.23** (103)	-.01 (103)	.06 (103)	.23** (103)	-.01 (103)	.06 (103)	-.15 (51)	-.19* (101)
Avg. speed 2.2	.02 (102)	-.02 (102)	.17 (102)	.16 (102)	-.13 (102)	.05 (102)	-.16 (50)	-.08 (100)
Avg. speed 3.1	-.01 (98)	-.10 (98)	.19* (98)	-.04 (98)	-.05 (98)	.03 (98)	-.09 (50)	-.01 (96)
Avg. speed 3.2	-.05 (100)	-.02 (100)	.09 (100)	.12 (100)	.02 (100)	.08 (100)	-.16 (50)	.02 (98)
Dead ends count 1.2	-.04 (102)	-.02 (102)	-.09 (102)	-.08 (102)	-.05 (102)	.11 (102)	.04 (51)	.06 (100)
Dead ends count 2.1	-.02 (103)	-.29*** (103)	-.17 (103)	.04 (103)	-.21** (103)	-.24** (103)	.00 (51)	-.23** (101)
Dead ends count 2.2	-.16 (102)	-.19* (102)	.03 (102)	.19* (102)	-.17* (102)	-.07 (102)	-.17 (50)	-.17* (100)
Dead ends count 3.1	-.23** (98)	-.22** (98)	-.20* (98)	.01 (98)	.01 (98)	-.12 (98)	.03 (50)	-.23** (96)
Dead ends count 3.2	-.02 (99)	-.25** (99)	.00 (99)	.11 (99)	-.03 (99)	-.07 (99)	-.21 (50)	-.09 (97)

Note. * $p < .1$, ** $p < .05$, *** $p < .01$.

Table 5.4 Rotated factor loadings resulting from the principal factor analysis

	Distance factor	Uniqueness		Inactive time factor	Uniqueness
Distance 1.1	0.56	0.68	.	.	.
Distance 1.2	0.50	0.75	Inactive time 1.2	0.02	0.99
Distance 2.1	0.28	0.92	Inactive time 2.1	0.46	0.79
Distance 2.2	0.31	0.91	Inactive time 2.2	0.35	0.88
Distance 3.1	0.40	0.84	Inactive time 3.1	0.53	0.72
Distance 3.2	0.44	0.80	Inactive time 3.2	0.28	0.92
	Average speed factor	Uniqueness		Dead ends count factor	Uniqueness
Avg. speed 1.1	0.37	0.86	.	.	.
Avg. speed 1.2	0.43	0.82	Dead end count 1.2	0.13	0.98
Avg. speed 2.1	0.75	0.44	Dead end count 2.1	0.64	0.60
Avg. speed 2.2	0.64	0.60	Dead end count 2.2	0.55	0.70
Avg. speed 3.1	0.47	0.78	Dead end count 3.1	0.52	0.73
Avg. speed 3.2	0.43	0.82	Dead end count 3.2	0.57	0.67

5.3.3 Factor Analyses

Four principal factor analyses were performed to retrieve the four factors for the six distance variables, the six inactive time variables, the six average speed variables, and the six dead ends variables each. Orthogonal Varimax rotation was utilized. The Kaiser-Meyer-Olkin measures of sampling adequacy were .61, .53, .64, and .71, respectively, indicating that the sampling for all groups of variables was above the acceptable limit of .5 (Field, 2009). Moreover, Bartlett's test of sphericity was significant for all groups of variables, with $\chi^2(15) = 51.75$, $p < .01$ for distance, $\chi^2(10) = 27.07$, $p < .01$ for inactive time, $\chi^2(15) = 107.70$, $p < .01$ for average speed and $\chi^2(10) = 60.67$, $p < .01$ for dead ends count. Cronbach's alpha was .56,

.36, .67 and .60 respectively. Table 5.4 displays the factor loadings and uniqueness of each variable contributing to the factor after rotation.

Table 5.5 Pearson's pairwise correlations between the EF test battery and the factors retrieved from the four CM variables factor analysis.

EF tasks	Factors							
	Distance factor (N)		Inactive time factor (N)		Average speed factor (N)		Dead ends count factor (N)	
ToL	-.26**	(94)	-.06	(96)	.13	(94)	-.14	(96)
RCPM	-.29***	(94)	-.25**	(96)	-.02	(94)	-.33**	(96)
FPT	-.15	(94)	-.31***	(96)	.11	(94)	-.12	(96)
VFT	.02	(94)	.19*	(96)	.22**	(94)	.13	(96)
DSfw	-.04	(94)	-.10	(96)	-.06	(94)	-.15	(96)
DSbw	-.14	(94)	-.23**	(96)	.09	(94)	-.17*	(96)
NGA	-.23	(47)	.16	(48)	-.19	(47)	-.13	(48)
PM	-.04	(92)	-.03	(94)	-.13	(92)	-.24**	(94)

Note. * $p < .1$, ** $p < .05$, *** $p < .01$.

Based on these factor loadings, the factor was extracted using the standard regression method. The resulting factors were correlated to the EF test battery using Pearson's pairwise correlation. We observed that the distance factor had a strong negative correlation with ToL and RCPM. The inactive time factor had significant, negative correlations with RCPM, FPT and DSbw. The average speed factor had a significant, positive correlation with the VFT. Lastly, the factor for the count of dead ends had significant, negative correlations with the RCPM and PM.

5.4 Discussion

In this paper, we aimed to investigate whether embedded sensor technology in a toy could yield understanding of EF in kindergarteners. Measuring EF in kindergarteners is important, as EF is related to numerous life outcomes (Diamond, 2013). Performing traditional EF measurements in kindergarteners requires a lot of time and expertise, and newer measurements that are easier to administer are performed on computers or tablets. In this study, we wanted to

perform measures using a real-life toy or object. For this purpose, we designed CM, which collects data on how children move a ball through a maze using a magnetic pen on a glass-looking surface. From the data collected by CM, we extracted variables on processes of solving the assignments, namely the distance they needed to solve mazes, the time they stood still to think or plan (i.e. inactive time), their average speed of moving through the maze and the number of dead ends they entered. These factors were used to relate to outcomes on a variety of existing EF tasks.

5.4.1 Interpretation of the Results

Correlational analyses revealed that significant correlations between the EF measures and CM measures tend to be grouped per variable category of CM. For example, the distance CM variables had two negative, significant correlations with ToL as well as with RCPM. We combined the six variables of each CM category into separate factors using factor analysis. This should be kept in mind when interpreting the results, as the internal consistency was not fully satisfactory. Furthermore, we also noticed a high uniqueness of each variable contributing to the factors. Other criteria for factor analysis were satisfactory, such as the Kaiser-Meyer-Olkin measures of sampling adequacy and Bartlett's test of sphericity.

We will first discuss the results of the correlational analysis between the CM factors and the EF tests with respect to convergent validity (i.e. degree to which relations that are to be expected are found). Significant, negative correlations found between the CM distance factor and Tower of London (ToL) as well as Raven CPM (RCPM) could indicate that CM distance measures visuospatial abilities, since RCPM has been associated with visuospatial reasoning and ToL with visuospatial memory. They also overlap regarding problem-solving ability or reasoning. The direction of the relationship indicates that children with higher visuospatial abilities and problem-solving need less distance in CM to solve the mazes. Nevertheless, we might have expected a relation with outcomes on paper mazes (PM) as well, as this task also measures visuospatial abilities, as well as with the Five-Point Test (FPT) as this task measures problem solving as well.

The CM factor for inactive time correlated with several EF tasks in a negative way, meaning that higher EF task performance is associated with lower inactive times on CM. It is not immediately clear why the relations are grouped in such a way for inactive time, as the described cognitive processes for RCPM, FPT and DSbw have limited overlap. At the same

time, this can be explained by the low Cronbach's alpha of this factor, potentially yielding unreliable correlations. For DSbw, a clear working memory component is needed, which is related to short-term memory use for performing the FPT as discussed earlier. Furthermore, perseverations on the FPT have been shown to be linked to verbal working memory capacity in the DSbw in adults (Goebel, Fischer, Ferstl, & Mehdorn, 2009). For RCPM however, a limited memory component is needed for solving the items.

The CM factor average speed has a significant positive correlation with VFT, which is a task that relies on verbal memory and language comprehension. The only other tasks that require a verbal component (memory or comprehension) are the digit span tasks, for which no correlations were found. VFT also has a very explicit time measurement aspect (i.e. children get one minute time). Nevertheless, we would then have expected a relationship between the factor for average speed and FPT as well, since this task also has an explicit time measurement aspect. It can furthermore be argued that general intelligence is underlying to the performance of the children on most tasks. This finding is therefore difficult to interpret.

Lastly, the CM factor for dead ends count had significant, negative correlations with RCPM and PM. These relations can indicate that this factor is related again to visuospatial abilities. Furthermore, PM has also been described to be an indicator of planning ability, but we see no relationship with ToL, which is another relevant task that measures planning. To conclude, regarding convergent validity, it seems that there is support for CM measuring EF. Particularly visuospatial abilities, such as visuospatial memory and reasoning, as well as problem-solving and a working memory component are related. No other relations have been found consistently to EF. This can potentially be explained by the reliability and validity of the EF tasks included as well as the reliability of the CM measures.

With respect to discriminant validity (i.e. the lack of relations that are not expected), the following can be concluded. There were two EF tasks that displayed no relation to the CM outcomes, which are NGA and DSfw. These tasks are both said to measure attentional processes specifically, and NGA further also measures inhibition and working memory. Due to the nature of the task in CM, it is somewhat surprising no relation is found with NGA. Nevertheless, it could be that CM does not pick up on attentional processes that are measured by both these EF tasks, though it can be argued that attention is used to some extent in the performance of all tasks in this study. We also notice that RCPM correlates with three out of

the four CM factors. This could be due to the fact that RCPM has been demonstrated to correlate strongly with non-verbal aspects of *g* or general intelligence (Raven et al., 1998), but of course, performance on the other EF tasks also depends upon general abilities. We can conclude that there are no relations found that were not expected.

It should be noted that PM, the task that most closely resembles the nature of the CM task, has only one significant correlation with CM, namely with the count of dead ends. We would have expected stronger relations between these two tests. Perhaps the testing medium or differences in the difficulty level of the mazes could explain this lack of findings. After all, the PM is a pencil-and-paper task, whereas CM requires the children to move a pen over a glass-looking surface to manipulate a ball. Also, PM consists of more items than the limited number of assignments made with the CM, which may limit the comparability.

5.4.2 Implications

The study highlights how a machine could be able to collect data on the performance of a child. Similar to a computer collecting information on task performance, a real toy could do the same. By letting children solve multiple mazes of different difficulty levels, information could be continuously collected. Once the child is familiarized with how the task works, different mazes could be made and information could be continuously saved on how children solve them. It could even be possible to send the information to a computer automatically, which could generate output via the use of predetermined algorithms about the EF of the child to an examiner. This study provides insights into the possibilities of using embedded sensor technologies in toys in testing the EF of kindergarteners. By combining different pieces of information collected by CM on the processes of solving mazes, we can relate the resulting factors to outcomes on EF tasks regarding visuospatial abilities (general but also visuospatial reasoning and memory), (problem) solving and to some extent working memory.

It is important to consider the potential influence of the testing medium on the outcomes. Csapó, Molnár and Nagy (2014) compared the performance of children on a computerized task to more traditional testing in terms of outcomes relating to school readiness and early development performed by teachers in a face-to-face setting. The test performance of children was lower in the computerized condition, which the authors linked to potential leniency when tested face-to-face as compared to automatic and computer-based assessment. They also offered potential difficulties using a computer as an explanation, but the authors did

not further specify. To our knowledge, there is no literature available that specifically deals with this topic with regard to toys. One may wonder what the influence could be that the test is presented to the children looking like a toy.

5.4.3 Limitations and Recommendations

Due to the design of the maze and the way the data were collected, some potentially interesting features of task performance can be considered for future research and toy development. For example, the time a child spends thinking right after the provision of the instruction of each assignment is not measured. It may be interesting to include these features in future measurements by adjustment of the protocol. It could also be interesting to include video images of a child while s/he is solving the assignments to be able to include off-task behaviour. Periods of non-movement could then be analyzed and coded (more) meaningfully. This could also be done using eye movement tracking. Furthermore, the design of our study included one toy. It would be interesting to include multiple toys that collect data on the EF of the child. This would allow the inclusion of more diverse data on the child and lead to better predictions of EFs and a more comprehensive view. Lastly, there were some technical difficulties with the measurements with the digital maze, such as lost connection or problems with the sensors, leading to a slightly smaller sample size than initially intended. More rigorous testing is needed to avoid issues in the future.

Our study focused on children of a specific age group. We recommend future research on the validity of CM to include multiple age groups. If CM would include more mazes to solve, the reliability of the factors could potentially increase, allowing more firm interpretations of the correlations. Based on the work of Heo, Kim and Faith (2015), we calculated that in order to reach a Cronbach's alpha of .7 for the four CM factors, that 11 distance items would have been needed for the distance factor, 21 items would have been needed for the inactive time factor, seven items for the average speed factor and eight items for the dead ends count factor.

Finally, the success rate of completing the maze patterns was also quite high in our study. Perhaps if more difficult maze patterns would have been included, valuable information on success versus fail could be included. We included several mazes of different difficulty levels. If more mazes are included of more similar difficulty levels, a factor analysis could be

performed with lower uniqueness of the individual variables. Possibly the correlations with the EF tasks could have been higher as well.

Our findings are very relevant for researchers working on the assessment of EF in kindergarteners. The results have implications for the use of smart toys and how they could be used to benefit assessment practices. With sufficient research, reliability of the measures could be increased and a more transparent picture could be painted if more diverse measures are included to compare to CM performance. For example, to link CM outcomes more strongly with other measures, comparisons could be made to questionnaires, systematic observations or even school performance. In the long run, the conclusions of this study can be useful to clinicians or educators as well, specifically regarding the ease of assessment in young children, provided there is sufficient validation and understanding of CM.

5.5 References

- Ardila, A., Ostrosky-Solís, F., & Bernal, B. (2006). Cognitive testing toward the future: The example of Semantic Verbal Fluency (ANIMALS). *International Journal of Psychology, 41*(5), 324-332.
- Ardila, A., Rosselli, M., & Rosas, P. (1989). Neuropsychological assessment in illiterates: Visuospatial and memory abilities. *Brain and Cognition, 11*(2), 147-166.
- Banken, J. A. (1985). Clinical utility of considering Digits Forward and Digits Backward as separate components of the wechsler adult intelligence Scale-Revised. *Journal of Clinical Psychology, 41*(5), 686-691.
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development, 81*(6), 1641-1660. doi:10.1111/j.1467-8624.2010.01499.x
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x
- Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning disabilities, 43*(6), 541-552.
- Brown, T. T., & Jernigan, T. L. (2012). Brain development during the preschool years. *Neuropsychology Review, 22*(4), 313-333. doi:10.1007/s11065-012-9214-1
- 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.
- Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616.
- 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.
- Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biology, 10*(3), e1001293.

- Criaud, M., & Boulinguez, P. J. (2013). Have we been asking the right questions when assessing response inhibition in go/no-go tasks with fMRI? A meta-analysis and critical review. *Neuroscience and Biobehavioral Reviews*, *37*(1), 11-23.
- Csapó, B., Molnár, G., & Nagy, J. (2014). Computer-based assessment of school readiness and early reasoning. *Journal of Educational Psychology*, *106*(3), 639-650.
doi:10.1037/a0035756
- Daigneault, S., Braun, C. M., & Whitaker, H. A. (1992). Early effects of normal aging on perseverative and non-perseverative prefrontal measures. *Developmental Neuropsychology*, *8*(1), 99-114.
- Di Lieto, M. C., Inguaggiato, E., Castro, E., Cecchi, F., Cioni, G., Dell’Omo, M., . . . Sgandurra, G. (2017). Educational robotics intervention on executive functions in preschool children: a pilot study. *Computers in Human Behavior*, *71*, 16-23.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168.
- Field, A. (2009). *Discovering statistics using SPSS*. California: Sage Publications.
- Goebel, S., Fischer, R., Ferstl, R., & Mehdorn, H. M. (2009). Normative data and psychometric properties for qualitative and quantitative scoring criteria of the Five-point Test. *The Clinical Neuropsychologist*, *23*(4), 675-690.
- Harrison, J. E., Buxton, P., Husain, M., & Wise, R. (2000). Short test of semantic and phonological fluency: Normal performance, validity and test-retest reliability. *British Journal of Clinical Psychology*, *39*(2), 181-191.
- Heo, M., Kim, N., & Faith M.S. (2015). Statistical power as a function of Cronbach alpha of instrument questionnaire items. *BMC Medical Research Methodology*, *15*, 1-9.
doi:10.1186/s12874-015-0070-6
- Johnstone, S. J., Dimoska, A., Smith, J. L., Barry, R. J., Pleffer, C. B., Chiswick, D., & Clarke, A. R. (2007). The development of stop-signal and Go/Nogo response inhibition in children aged 7–12 years: performance and event-related potential indices. *International Journal of Psychophysiology*, *63*(1), 25-38.
- Jonkman, L. M. (2006). The development of preparation, conflict monitoring and inhibition from early childhood to young adulthood; a Go/Nogo ERP study. *Brain research*, *1097*(1), 181-193.
- Józsa, K., Caplovitz-Barrett, K., & Morgan, G. A. (2017). Game-like tablet assessment of approaches to learning: Assessing mastery motivation and executive functions. *Electronic Journal of Research in Educational Psychology*, *15*(3), 665-695.

- Kaller, C. P., Unterrainer, J. M., & Stahl, C. (2012). Assessing planning ability with the Tower of London task: Psychometric properties of a structurally balanced problem set. *Psychological Assessment, 24*(1), 46.
- Kaufman, A. S. (1990). The WPPSI-R: You can't judge a test by its colors. *Journal of School Psychology, 28*(4), 387-394.
- 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., Resing, W.C.M. & Verhaeghe, P. (2002). *WISC-III NL. Handleiding*. London: The Psychological Corporation.
- Langenecker, S. A., Zubieta, J.-K., Young, E. A., Akil, H., & Nielson, K. A. (2007). A task to manipulate attentional load, set-shifting, and inhibitory control: Convergent validity and test-retest reliability of the Parametric Go/No-Go Test. *Journal of Clinical and Experimental Neuropsychology, 29*(8), 842-853.
- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology, 21*(1), 59-80.
- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2004). *Neuropsychological assessment*. New York, US: Oxford University Press.
- Lunt, L., Bramham, J., Morris, R. G., Bullock, P. R., Selway, R. P., Xenitidis, K., & David, A. S. (2012). Prefrontal cortex dysfunction and 'jumping to conclusions': bias or deficit? *Journal of Neuropsychology, 6*, 65-78.
- Miller, H., Barnes, J., & Beaver, K. M. (2011). Self-control and health outcomes in a nationally representative sample. *American journal of health behavior, 35*(1), 15-27.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, 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.
- Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. *Child development at the intersection of emotion and cognition, 1*, 203-224.

- Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory*, 7(2), 209-231.
- 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, Texas: 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.
- Riva, D., Nichelli, F., & Devoti, M. (2000). Developmental aspects of verbal fluency and confrontation naming in children. *Brain and Language*, 71(2), 267-284.
- Shallice, T. (1982). Specific impairments of planning. *Phil. Trans. R. Soc. Lond. B*, 298(1089), 199-209.
- Strauss, E., Sherman, E. M., & Spreen, O. (2006). *A compendium of neuropsychological tests*. New York: Oxford University Press.
- Tombaugh, T. N., Kozak, J., & Rees, L. (1999). Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Archives of Clinical Neuropsychology*, 14(2), 167-177.
- Tsuchiya, E., Oki, J., Yahara, N., & Fujieda, K. J. B. (2005). Computerized version of the Wisconsin card sorting test in children with high-functioning autistic disorder or attention-deficit/hyperactivity disorder. *Brain and Development*, 27(3), 233-236.
- 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.
- 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 Experimental Neuropsychology*, 26(6), 846-856.
- Vander Steene, G., & Bos, A. (1997). WPPSI-R: Wechsler preschool and primary scale of intelligence (Dutch-Flemish version). Lisse: Swets & Zeitlinger.

- Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology, 140*, 38-55.
- 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.
- Whiteside, D. M., Kealey, T., Semla, M., Luu, H., Rice, L., Basso, M. R., & Roper, B. (2016). Verbal fluency: Language or executive function measure? *Applied Neuropsychology: Adult, 23*(1), 29-34.



Chapter 6

Free play and executive functioning:

What can Clever Maze reveal?

This chapter is based on joint work with Seethu Christopher^a, Karien Coppens^b,
Petra Hurks^c, Rico Möckel^a and Lex Borghans^b.

^a Department of Data Science and Knowledge Engineering, Faculty of Science
and Engineering, Maastricht University

^b Department of Macro, International and Labour Economics, School of Business
and Economics, Maastricht University

^c Department of Neuropsychology and Psychopharmacology, Faculty of
Psychology and Neuroscience, Maastricht University

6.1 Introduction

The importance of play for cognitive development in children has been established in research and is generally accepted by the public. Particularly before children enter formal primary education, play is an important mechanism via which children learn. Play can be very diverse, from fantasy to rolling around, and each type of play contributes to brain, social and cognitive development (Goldstein, 2012). Didactic programmes in kindergarten are therefore also aimed at teaching kindergarteners via play or game-like toys and approaches, in order to appeal to children. Most didactic materials designed for young children, therefore, embody this philosophy. For example, toys are designed to stimulate various aspects of cognitive development and within the average home and kindergarten classroom, ample toys are provided in such a way they provoke play and cognitive abilities. However, regarding assessments, materials and test procedures are often less appealing to kindergarteners. Yet, understanding cognitive ability from an early age onward is very important to track development and identify potential deficits.

Standardized tasks are the go-to tool to assess development or cognitive skills, but the availability of such tasks and kindergarteners' ability to cooperate in lengthy evaluations are limited (Isquith, Gioia, & Espy, 2004). The use of standardized tests has become widely implemented as a standard tool for understanding cognitive abilities from a very early age onward. However, research has shown that test-related stress or test anxiety can take place in children when performing such tests (Segool, Carlson, Goforth, Von Der Embse, & Barterian, 2013; von der Embse, Jester, Roy, & Post, 2017). These effects are much harder to study with children that are even younger than elementary school age since self-report measures are difficult to use in kindergarteners. However, the potential effects of standardized testing in kindergarten are likely to be similar. It seems there is friction between the skills that are being promoted and the means of assessment that are available for understanding them, as educators voice the discrepancy between these materials and the reality (Werkgroep en Steungroep Kleuteronderwijs, 2013).

Since studies have shown kindergarteners spend one of the largest portions of their time on play (Hofferth & Sandberg, 2001), it is worth investigating how assessments could be more aligned with play. This line of thought implies moving away from the standardized test environment, toward an assessment that takes place in a naturalistic setting. In this study, we

investigate whether information can be collected from the play of kindergarteners, in an objective manner. We will first highlight the role of play and how it can be measured, followed by the role technology can play in this. Next, we will zoom in on cognitive abilities and the framework that is used in this study. Lastly, the aim of the current study will be explained, where innovative and technological means will provide answers to the main research question.

6.1.1 The Role and Assessment of Play

Current existing, cognitive tests for kindergarteners are mostly administered in a standardized test environment, where a trained examiner collects data from the children by noting observations and results. Usually, such tests take place in a one-to-one setting. Often these tasks also require extensive scoring after the test. The process requires therefore high time and expertise investment from the examiner. For example, children imitate hand motions of the examiner in Luria's Hand Game as a measure of inhibitory control (Hughes, 1998). Another example is a game where children have to perform a motor sequence as fast as they could on a musical keyboard until the examiner says stop. Observers score the children on the number of sequences correctly completed within 10 seconds as a measure of planning ability (Carlson & Moses, 2001; Welsh, Pennington, & Groisser, 1991).

In a similar fashion, a researcher or educator can observe a child's free playing behaviours and make notes on predetermined characteristics of the play. This type of measurement already encompasses a more natural process from the perspective of the child. The Play Observation Scale (POS) (Rubin, 2001) has often been used to study play behaviour in group context and was later also used to study non-social forms of play. The Preschool Play Behavior Scale is another example which teachers can use to assess social and solitary play (Coplan & Rubin, 1998), which leans more toward a survey-type of data collection. In the POS, various play behaviours are coded for at least 15 minutes, such as solitary, parallel or group play, functional, constructive or dramatic play, game-with-rules, exploratory, etc. Furthermore, surveys and questionnaires have been developed that can be filled in by teachers or parents. The BRIEF (Behavioral Rating Inventory of Executive Functioning) questionnaire is a good example which is designed to evaluate (young children's) executive functions (EF) based on everyday behaviours. It also includes items on play behaviour when assessing a younger age group (Isquith et al., 2004). As in the above-mentioned examples, the examiner has to interpret the scores on the BRIEF. The examiner is usually an expert who is extensively

trained in studying child development. Given the large time and expertise investments needed, large-scale implementation of such assessments are therefore difficult to realize.

6.1.2 The Role of Technology

Currently, more means have become available to collect information on behaviour besides observation or survey methods thanks to technological advances. Sensor technology is already employed in some toys to create ‘smart toys’, where information on the interaction between the toy containing the sensors (e.g. robot) and child is utilized for a variety of purposes. Such toys have been used to great advantages in certain fields of developmental stimulation. As an example of how smart toys are currently used in research, Belpaeme et al. (2015) and Vogt et al. (2019) studied the use of social robots to enhance second language learning in early childhood but found limited added value compared to a tablet. Ornelas Barajas, Al Osman and Shirmohammadi (2017) utilized smart building blocks with embedded electronic modules which provide visual feedback via a computing device as a serious game for children with an Autism Spectrum Disorder. These studies are mostly aimed at improving certain skills. Interestingly, these toys could also be used to collect data on the current skill level of children. If developed properly, these toys could make assessment less time consuming and complicated, whilst maintaining a high level of precision and objectiveness. So rather than using smart toys to increase certain skills in an intervention setting, the smart toy itself could be a measurement tool.

If the toy is smart enough to collect data of children interacting or playing with it, this could bring numerous additional advantages. The first would be the ease with which data are collected, as sensors could store the information in the tool itself or send it to a computer. The second is accuracy, as toys can be designed to collect any type of data that are not subjected to the administrator’s level of expertise. Third, continuous measurement could be possible, as the toy could simply be placed inside a classroom for continuous use by the children without continuous one-to-one supervision. It would be relatively easy to collect data on skill growth, calculate differences between children, and track deficits, for example.

6.1.3 Cognitive Skills and Assessments in Kindergarten

It is important to gain an understanding of which cognitive abilities can be measured in the kindergarten age group and how children are developing on these skills. It is well understood

in this body of research that cognitive abilities are undergoing heavy construction before children enter formal primary education. A specific and important type of cognitive abilities are EF, which are cognitive skills that regulate everyday behaviour (Best & Miller, 2010). EF consist of three basic cognitive components, namely working memory, inhibitory control, and cognitive shifting (Diamond, 2013; Lehto et al., 2003; Miyake et al., 2000). These basic EF processes allow for higher-order EF processes such as reasoning, problem-solving and planning (Collins & Koechlin, 2012; Lunt et al., 2012). EF have been shown to be important predictors of important life outcomes such as physical health (Miller, Barnes, & Beaver, 2011), quality of life (Brown & Landgraf, 2010), school success (Borella, Carretti, & Pelegrina, 2010), school readiness (Blair & Razza, 2007; Morrison, Ponitz, & McClelland, 2010) and job success (Bailey, 2007). While initial research on EF focused on studying executive dysfunction in atypical populations, such as children or adults with brain lesions, variation of the normal population has been shown to be meaningful and predictive of relevant outcomes such as academic achievement in more recent research (Carlson & Moses, 2001). Given the link to many life outcomes, gaining insight into EF early in life is therefore valuable. Early measurements of EF such as in kindergarten, even in normative populations, could provide opportunities to track and stimulate development.

6.1.4 The Current Study

The main question of this study is to what extent behaviour during free play situations can be linked to EF. We use a smart toy called Clever Maze (CM). CM looks like a regular wooden toy, with a ball trapped in a maze under a glass-looking surface (see Chapter 4 for detailed information). Children can manipulate the ball through the maze with by moving the pen over the surface of the maze. Different designs can be inserted into CM. In this study, kindergarteners are invited to play with the CM in an unstructured way. Unstructured play means that we will encourage them to play with the CM, but no predetermined goal or objective is provided to children. For instance, they do not have to go to certain areas within the maze and there is no time pressure. Several variables can be extracted from the CM on the playing behaviour of children. We relate their playing behaviour to their performance on standardized EF tests, as well as their parents' answers on the BRIEF questionnaire.

The inclusion of standardized, one-to-one EF tasks allows the comparison of CM outcomes to outcomes of the children from actual assessment situations. The comparison of

CM data to the BRIEF outcomes allows valuable insights into behaviour as measured by a questionnaire asking about everyday behaviour. As the play is not structured, a link may be present between everyday behaviour as measured with the BRIEF and variables measured by the unstructured play in CM. The results will be discussed in light of convergent and discriminant validity. As this is an exploratory study, no detailed hypotheses are formulated. We aim to provide a first step in answering the question of whether testing without explicit testing is possible, i.e. to collect meaningful data on spontaneous behaviour in young children. Results will be discussed in-depth in the discussion, both on the level of the existence of specific relationships between variables as well as linking our findings to the general question on whether measuring free play is a valid way to understand EF.

6.2 Method

6.2.1 Participants

Elementary schools in the Southern region of the Netherlands were invited to participate in the study. Information letters, consent forms and the BRIEF questionnaire were distributed among parents of children in group 2. Group 2 is the last year of kindergarten when children are on average 6 years old. We excluded children who had repeated a grade, but also children who did not have at least one parent who sufficiently understood Dutch since all information was provided in this language. Finally, we also did not include children with a diagnosed disorder such as a developmental or psychiatric disorder. The criteria were included in order to focus on the population of children within a normal range of development, rather than having developmental or language issues as a potential confounder in the analyses.

6.2.2 Procedures and Materials

For this study, the same testing protocol was used for the EF test battery and CM as in Chapter 5. The test included in the EF test battery were Raven Coloured Progressive Matrices (RCPM), Tower of London (ToL), Paper Mazes (PM), Five-Point Test (FPT), Verbal Fluency Test (VFT), Digitspan Forward and Backward (DSfw, DSbw), Go/No-Go (No-go accuracy was utilized, NGA abbreviated). Detailed descriptions of these tests can be found in Chapter 5, specifically section 5.2.2. The outcomes of the BRIEF questionnaire were also included. With CM, a unique maze design was used for the ‘play’-condition in this study. Detailed description of the functioning of CM can be found in Chapter 4 and 5.

Design of play condition in Clever Maze.

For this study, a play board was inserted into CM. The design of the play board can be seen in figure 6.1. The board consisted of an intricate design with four 'point of interest' (i.e. circular areas where in the assignment condition a star would be included), but for the purpose of free play. The design consisted of 18 dead ends that children could enter. At the start of the data collection, the ball would be placed in the centre of the maze and can be manipulated by the children via the use of a pen, as shown in figure 6.1.



Figure 6.9 Top picture shows the design of the play maze, as seen from the researcher's perspective. The bottom picture displays the pen children use to manipulate the ball inside the maze and button and feedback-LEDs on the side of the researcher.

Procedure play condition of the Clever Maze.

A detailed protocol was devised for the data collection with CM, which can be found in Appendix C. The instruction of the play condition was brief, stating that children would be allowed to play with a maze. The researcher asked the child if s/he knew what a maze is. Then the maze pattern was revealed. The researcher explained to the child that s/he could move the ball inside the maze using the pen and that they could play freely with the maze. A stopwatch was used to track how long children played, starting from the moment instruction was finished and children started to play. This served as insight to the researcher during testing and this was therefore not recorded, as CM is tracking time as well. If the child did not do anything, the researcher would encourage the child to try something. If the child was truly unable to move the ball or interact with the maze, measurement would be ceased after 90 seconds. If the child stopped playing after less than 2 minutes, the researcher indicated that the child still had time to play if s/he wanted. If the child played between 2-5 minutes, the researcher did nothing until (non-)verbal communication of the child clearly indicated s/he is done playing. If the child played longer than 5 minutes, the researcher would wrap up the play.

BRIEF.

We utilized the Dutch version of the BRIEF questionnaire for 5-17-year-olds (Huizinga & Smidts, 2009). The original American version was designed by Gioia, Isquith, Guy, and Kenworthy (2000). The BRIEF aims to map executive functioning skills based on the description of behaviour. For this study, we used the teacher form. The questionnaire contained 75 items divided across eight subscales, two general indices, a total score and two validity scores. For this study, we used the scores on the subscales, which each had a specific range of item scores: inhibition (10-30), flexibility (10-30), emotion regulation (9-27), initiative-taking (7-21), working memory (10-30), planning and organizing (10-30), order and neatness (7-21), behavioural evaluation (10-30). Two items explicitly asked about the play of the child, which are both a part of the 'initiative taking'-scale (i.e. item 4 and 16). The BRIEF was designed as a clinical scale, meaning that a higher score reflects more problematic behaviour and thus lower EF skills. Huizinga and Smidts (2009) reported the following psychometric properties for the Dutch BRIEF. For children aged 5-7 years old, the reliability of the raw scores on all subscales of the teacher form was between .75 and .96. Internal consistency, as measured using Cronbach's alpha, was at least .87 for all subscales. Test-retest reliability was satisfactory as

well. It has been shown to be a valid measure of everyday functioning (Gioia et al., 2000; Kenworthy, Yerys, Anthony, & Wallace, 2008; Toplak, Bucciarelli, Jain, & Tannock, 2008).

6.2.3 Data and Analyses

The collected data from the play phase were transformed into variables. The first was the total distance travelled expressed in cm for the whole time the child played. We also extracted the number of dead ends children entered during this time, how much time they spent in the dead ends in seconds and how much distance they travelled in the dead ends in seconds. Their average speed of moving was also included, which was an average value of the time between each collected data point. Lastly, the time they stood still in the maze was also included in seconds. Descriptive statistics provide insight into the behaviour of children within the play condition. Correlational analyses are used to study the relationship between the CM variables, the EF test battery outcomes and the BRIEF outcomes. We use Spearman's non-parametric correlation as it is possible non-linear relation could exist. For all analyses, we used variables where all outliers received a floor-and ceiling value of three standard deviations plus or minus the mean.

6.3 Results

The result section first provides descriptive statistics on the CM variables as well as the EF variables, followed by correlational analysis using Spearman's non-parametric correlation.

6.3.1 Descriptives

The sample consisted of 121 primary school children in the Limburg region of the Netherlands. 13 children were excluded due to absence on the days of testing or technical difficulties in CM. In total, 108 children were included in the analysis with a mean age of 72.49 months (6 years) and a standard deviation of 4.37 months. There were 54 boys and 54 girls in the sample. Table 6.1 shows descriptive information on the variables included in the analysis. We can observe from table 6.1 that there was sufficient range in the variables included for this study regarding the play condition.

Table 6.1 Descriptive information on the EF and CM variables.

EF tasks	Obs.	Mean	SD	Min	Max
ToL	108	13.09	2.47	5.07	18
RCPM	108	23.34	4.37	10.14	34
FPT	108	7.11	2.63	1.00	14.00
VFT	108	10.64	3.72	3.00	21.00
DSfw	108	6.12	1.76	3.00	11.00
DSbw	108	2.66	1.15	0.00	5.00
NGA	52	0.69	0.17	0.29	0.98
PM	106	10.49	2.49	6.00	18.05
CM					
Distance ^a	108	1297.03	506.73	165.84	2415.48
Time ^b	108	251.81	73.52	31.23	336.07
Number of dead ends	108	25.33	11.41	2.00	56.00
Distance in dead ends ^a	108	303.51	161.20	20.82	787.15
Time in dead ends ^b	108	53.48	23.47	5.68	118.35
Average speed	108	16.39	10.67	1.07	49.47
Inactive time ^b	108	20.06	16.98	0.00	71.13
BRIEF scales					
Inhibition	104	14.37	3.40	10	24.80
Flexibility	104	11.18	2.54	8	18
Emotion regulation	104	14.56	3.47	10	25.39
Initiative taking	104	10.93	2.47	8	18.74
Working memory	104	14.63	3.53	10	25.29
Planning	104	14.73	2.44	12	22.35
Order and neatness	105	9.15	2.11	6	15.52
Behavioral evaluation	105	11.94	3.02	8	21

Note. ^a Variables are expressed in centimetres. ^b Variables are expressed in seconds.

6.3.2 Correlational Analyses

Spearman's non-parametric correlation was performed to understand the relation between the variables extracted from CM, the tasks included in the EF test battery and the outcomes on the BRIEF questionnaire. The results of the correlations between the EF tasks and the Clever Maze are depicted in table 6.2. Table 6.3 depicts the relations between the Clever Maze and the BRIEF scales.

Table 6.2 Spearman’s non-parametric correlation between the CM variables and the outcomes on the EF tests.

CM variables	EF test scores							
	ToL (N)	RCPM (N)	FPT (N)	VFT (N)	DSfw (N)	DSbw (N)	NGA (N)	PM (N)
Distance ^a	.15 (108)	.03 (108)	.09 (108)	.11 (108)	.11 (108)	.14 (108)	-.15 (52)	-.15 (106)
Time ^b	.10 (108)	.00 (108)	.00 (108)	.10 (108)	.12 (108)	.19** (108)	-.17 (52)	-.15 (106)
Dead ends entered	-.07 (108)	-.02 (108)	-.03 (108)	.01 (108)	-.01 (108)	.07 (108)	-.05 (52)	-.17* (106)
Distance in dead ends ^a	-.16* (108)	-.07 (108)	-.01 (108)	.04 (108)	-.01 (108)	-.10 (108)	-.15 (52)	-.27*** (106)
Time in dead ends ^b	-.31*** (108)	-.16* (108)	-.13 (108)	-.08 (108)	-.12 (108)	-.11 (108)	-.21 (52)	-.25** (106)
Average speed	-.08 (108)	-.01 (108)	-.06 (108)	.14 (108)	.05 (108)	-.05 (108)	-.16 (52)	-.19* (106)
Inactive time ^b	.00 (108)	.06 (108)	-.20** (108)	-.05 (108)	-.02 (108)	.09 (108)	.26* (52)	-.02 (106)

Note. ^a Variables are expressed in centimetres. ^b Variables are expressed in seconds. * p < .1, ** p < .05, *** p < .01.

Table 6.3 Spearman’s non-parametric correlation between the CM variables and the BRIEF subscales. N=104

CM variables	BRIEF subscales							
	Inhibition	Flexibility	Emotion regulation	Initiative taking	Working memory	Planning and organizing	Order and neatness	Behavioural evaluation
Distance	-.13	-.09	-.07	-.25**	-.12	-.04	.01	-.14
Time	-.09	-.08	-.04	-.18*	-.14	-.08	.03	-.09
Dead ends entered	-.12	-.06	-.05	-.15	-.07	.05	.01	-.07
Distance in dead ends	-.08	-.05	-.02	-.20**	-.12	.01	.03	-.11
Time in dead ends	.02	.03	.05	-.07	.04	.17*	.14	.05
Average speed	-.11	-.22**	-.14	.00	-.10	-.10	.02	.00
Inactive time	.12	.01	.00	.06	.07	.03	.04	.04

Note. * p < .1, ** p < .05, *** p < .01.

Table 6.3 shows several significant, negative correlations between the CM variables and the outcomes on the EF tasks. Higher distance travelled in and longer time spent in dead ends is associated with lower performance on the PM. Spending more time in dead ends is also associated with a lower score on the ToL. Furthermore, a longer inactive time is associated with a lower score on the FPT. A significant positive correlation found was between a longer time spent playing in the maze and a higher score on the DSbw.

Table 6.4 displays several significant, negative correlations between parents' responses on the BRIEF questionnaire and their child's playing behaviour as recorded by CM. A higher score on the clinical subscale 'Initiative taking' is related to travelling less distance in CM while playing in general, as well as in dead ends. A higher score on the subscale 'Flexibility' is also associated with a lower speed of moving through CM while playing.

Table 6.4 Spearman's non-parametric correlation between the EF task outcomes and the BRIEF subscales.

CM variables	EF test scores							
	ToL (N)	RCPM (N)	FPT (N)	VFT (N)	DSfw (N)	DSbw (N)	NGA (N)	PM (N)
Inhibition	.03 (104)	-.04 (104)	-.08 (104)	.01 (104)	-.17 (104)	-.01 (104)	-.27* (48)	.01 (102)
Flexibility	-.05 (104)	.03 (104)	.04 (104)	-.06 (104)	-.15 (104)	.03 (104)	-.14 (48)	-.06 (102)
Emotion regulation	.05 (104)	.02 (104)	-.08 (104)	.00 (104)	-.13 (104)	.19* (104)	-.26* (48)	.09 (102)
Initiative taking	-.15 (104)	-.12 (104)	.11 (104)	-.09 (104)	-.13 (104)	-.05 (104)	-.26* (48)	.08 (102)
Working memory	.00 (104)	-.02 (104)	.02 (104)	-.08 (104)	-.17* (104)	.00 (104)	-.22 (48)	.12 (102)
Planning	-.08 (104)	-.13 (104)	-.02 (104)	-.06 (104)	-.20** (104)	-.06 (104)	-.21 (48)	.00 (102)
Order and neatness	-.01 (104)	.13 (104)	.06 (104)	.07 (104)	.07 (104)	.24** (104)	-.41*** (48)	.07 (102)
Behavioral evaluation	-.04 (104)	-.17* (104)	-.12 (104)	-.03 (104)	-.16* (104)	.00 (104)	-.31** (48)	.01 (102)

Note. * $p < .1$, ** $p < .05$, *** $p < .01$.

To deepen the understanding of the relations found, we also performed a correlational analysis between the outcomes on the EF tasks and the BRIEF questionnaire. Particularly since the correlations between CM and BRIEF revealed that the subscale 'Initiative taking' seems to hold relevant information, we wished to see if this relationship was also present between the

BRIEF outcomes and EF scores. The results are displayed in table 6.4, where we observed that particularly the subscale ‘Order and neatness’ had significant relations to the EF task outcomes. A higher score on this clinical scale was associated with a lower score on the NGA and a higher score on the DSbw. Furthermore, the clinical subscale ‘Planning’ had a significant, negative correlation to DSfw. However, no relations with the subscales ‘Initiative taking’ or ‘Flexibility’ was observed.

6.4 Discussion

The importance of play has been widely recognized in education and pedagogical approaches. Studies have shown that play has a crucial role in child cognitive development. Nevertheless, assessment tools for understanding cognitive abilities often do not use play but require standardized, potentially strenuous test environments. In this study, we utilized information from children’s play that was collected by embedded sensor technology in a toy called CM, to explore whether information on cognitive ability can be retrieved from their play. CM continuously collected information on the playing behaviour of the children, i.e. how they moved the ball through the maze while no game or goal was provided. The play behaviour was related to EF via EF tasks and an EF-related parental questionnaire.

We extracted several variables from CM and correlated them to the children’s outcomes on several EF tasks and parent’s answers on the BRIEF questionnaire regarding their child’s EF. We observed that performance on the PM, which measures visuospatial abilities and planning, had negative correlations with information collected from the dead ends, particularly with distance and time spent in dead ends. Regarding dead ends variables, we also observed that a longer time spent in dead ends is significantly related to lower performance on the ToL, a task designed to measure planning and problem solving, as well as working memory. We can therefore conclude that variables collected on dead ends, particularly the time spent in them, provides insight into planning ability of children in the sense that spending more time and travelling more distance in dead ends is related to lower planning abilities. ToL and PM also overlap in measuring visuospatial abilities, with ToL measuring visuospatial memory specifically. Furthermore, the total time spent playing related significantly to performance on DSbw in a positive manner, a task measuring working memory. The total time not moving in CM is correlated negatively with the FPT, a task that measures problem solving, memory and inhibition abilities. This relationship again highlights the importance of problem-solving,

which is also measured by the ToL. Furthermore, the relationship found with ToL, FPT and DSbw also highlights working memory, indicating that CM is also picking up on memory aspects such as working memory and visuospatial memory with diverse CM measures. Regarding the relation between the BRIEF variables and CM, we observed that the scale 'initiative taking' correlated in a negative manner to the distance travelled in CM, both inside and outside dead ends. This is also the sub-scale of the BRIEF that includes two items on play behaviour of the child. The average speed of moving in CM was negatively related to the subscale 'flexibility'. This means that children who have more problems with taking initiative travel less distance in the maze while playing. Similarly, children who are experiencing problems with being flexible move slower through the maze.

In terms of interpreting the findings further, we examine the convergent validity. Taking together the findings described above, it seems when letting children play with CM in an unstructured way, CM can pick up on the children's planning abilities, problem-solving and working memory skills as measured by EF tasks. We also see a relationship with everyday behaviour as measured by the BRIEF in terms of initiative-taking and flexibility. The scale on initiative-taking included items (i.e. questions in the BRIEF) that rated the behaviour of the children in the context of play. Therefore, this relation between play behaviour in CM and this subscale is particularly relevant for the validity of CM. Regarding discriminant validity, we did not see any relations that were surprising or unexpected. The relations between CM and EF task outcomes as well as between CM and the BRIEF cannot be explained by relations between the EF task outcomes and BRIEF outcomes.

It should be noted that the correlations of CM with the BRIEF and the EF tasks were not particularly high. This can be explained by the fact that the children were not asked to perform a specific task in order to obtain a specific goal, as is often the case with the EF tasks and also aligned with the definition of EF (i.e. goal-directed, self-regulated behaviour). In a situation where natural behaviour was allowed to unfold, potential relationships with daily EF as reported in the BRIEF would have been plausible. Thus, we should have seen stronger relations between CM and BRIEF than between CM and EF tasks, which is not the case. However, the general procedure with CM as in this study may still feel artificial to the children and they may well still treat it as a testing situation. It could also be argued that the behaviours (i.e. items) reported in the BRIEF pertain to more of a collection of complex behaviours, which are not directly relatable to the more specific CM variables. The finding of low correlations between

the BRIEF and actual EF tasks is not uncommon (Anderson, Anderson, Northam, Jacobs, & Mikiewicz, 2002).

6.4.1 Limitations and Recommendations

A few remarks about our study should be provided. We let children play in an unstructured way with CM. However, there is still a researcher with the child observing the data collection, and there is only one toy to play with, namely CM. The set-up, compared to free play in a classroom under the supervision of a teacher, may feel a bit unnatural to children, which could potentially bias their behaviour in some way. In addition, we had to exclude several children from the sample due to measurement error, for example missing Bluetooth connection with CM. Furthermore, more work is needed to understand the exact constructs that are measured using a free play condition in the CM. We would, therefore, make several recommendations for more research.

Regarding CM it would be useful to study more age groups. The materials could be varied more by inserting different patterns with differing levels of simplistic and complex maze designs. It would also be highly interesting to apply the same line of thought to other toys that are more dynamic. For simplicity, we chose a static toy as a starting point. But by including toys that can be used in different ways, such as dolls or blocks, valuable insights could be gathered. If the mechanism is sufficiently tested, the setting could also be changed from a one-to-one to other settings that could make the assessment less artificial. Utilization of multiple toys could provide a comprehensive and continuous view on child EF, without strenuous assessments.

6.4.2 Implications of the findings

Toy-based tests could play an important role in making assessments more in line with the interests of children and their daily activities. It may be possible to uncover more potential in “testing without testing” by extensively studying what information can be uncovered from free play situations using sensor technology. It requires rigorous research in types of materials, toys and sensors that are suitable and usable in such target groups. The validity and reliability of said measures should always be monitored carefully. In the future, when a solid understanding of the type of tools, for which age group and for which questions specifically is obtained, implementation in a wider net could be very beneficial for the assessment experience of the

child. The experience shifts from a standardized testing environment to a playful moment, where play with a toy without wires is possible. This ease of assessment is also obvious for the examiner. Learning to use the toy as an assessment tool would be easy in comparison to many existing tests because of all the reasons mentioned earlier and no detailed, high-level training would be required for the examiner. The time and expertise investment that is required to perform cognitive assessments on kindergarteners could be reduced significantly. After all, the child could be able to play with the toy without up-close supervision of the examiner, meaning no specific instructions or registrations are required other than a child identifier. The data can be exported easily to a computer. Ease of assessment would be even further increased if the toy would be able to recognize which child is playing with the toy and send the data to a computer autonomously. If a server is developed where the data are stored and analysed via pre-made algorithms, the examiner can simply log in to the server and view the results of the child's play. Such a platform for data could also provide opportunities to not only provide simple reports per play but an overview of play over time. The type of information collected can be extended and explored further. Perhaps information on various sub-processes of playing in the maze yield interesting information. In the current study, we used outputs that closely resemble assignment-like outputs: how much distance they travelled, dead ends entered etc. Detailed information on the play could be analysed with state-of-the-art analysis techniques such as machine learning, potentially providing very comprehensive insights into EF.

The insights of our study and of potential future studies in the same line provide insights into the possibilities of using technology to the advancement of understanding cognitive abilities, particularly in target-groups that are more difficult to assess. Not only for young children, but also for the elderly this could potentially be interesting. For now, our insights and findings are relevant to the scientific world, but the mechanism and implications have to be studied in-depth. Future studies and applications can broaden the potential for CM and similar tools. CM is, of course, one toy that appeals to specific EF. We imagine that including other 'smart toys' with similar data-collecting properties as CM could provide a more comprehensive view of EF.

6.5 References

- Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships between cognitive and behavioral measures of executive function in children with brain disease. *Child Neuropsychology*, *8*(4), 231-240.
- Bailey, C. E. (2007). Cognitive accuracy and intelligent executive function in the brain and in business. *Annals of the New York Academy of Sciences*, *1118*(1), 122-141.
- Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., . . . Göksun, T. (2015). *L2TOR-second language tutoring using social robots*. Retrieved from <https://pub.uni-bielefeld.de/download/2900267/2900268/BelpaemeEA2015.pdf>
- Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641-1660. doi:10.1111/j.1467-8624.2010.01499.x
- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x
- Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning disabilities*, *43*(6), 541-552.
- Brown, T. E., & Landgraf, J. M. (2010). Improvements in executive function correlate with enhanced performance and functioning and health-related quality of life: evidence from 2 large, double-blind, randomized, placebo-controlled trials in ADHD. *Postgraduate Medicine*, *122*(5), 42-51.
- Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032-1053.
- Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PloS Biology*, *10*(3), e1001293.
- Coplan, R. J., & Rubin, K. H. (1998). Exploring and assessing nonsocial play in the preschool: The development and validation of the Preschool Play Behavior Scale. *Social Development*, *7*(1), 72-91.
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135-168.
- Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *BRIEF: Behavior rating inventory of executive function*. Odessa, Florida: Psychological Assessment Resources.

- 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>
- Hofferth, S. L., & Sandberg, J. F. (2001). How American children spend their time. *Journal of Marriage and Family*, *63*(2), 295-308.
- Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology*, *16*(2), 233-253.
- Huizinga, M., & Smidts, D. (2009). BRIEF Executieve Functies Gedragsvragenlijst: Handleiding [Dutch Manual BRIEF Behavior Rating Inventory of Executive Function]. *Amsterdam: Hogrefe Uitgevers*.
- Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, *26*(1), 403-422.
- Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding executive control in autism spectrum disorders in the lab and in the real world. *Neuropsychology Review*, *18*(4), 320-338.
- Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59-80.
- Lunt, L., Bramham, J., Morris, R. G., Bullock, P. R., Selway, R. P., Xenitidis, K., & David, A. S. (2012). Prefrontal cortex dysfunction and 'jumping to conclusions': bias or deficit? *Journal of Neuropsychology*, *6*(1), 65-78.
- Miller, H., Barnes, J., & Beaver, K. M. (2011). Self-control and health outcomes in a nationally representative sample. *American journal of health behavior*, *35*(1), 15-27.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, 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.
- Morrison, F. J., Ponitz, C. C., & McClelland, M. M. (2010). Self-regulation and academic achievement in the transition to school. *Child development at the intersection of emotion and cognition*, *1*, 203-224.
- Ornelas Barajas, A., Al Osman, H., & Shirmohammadi, S. (2017). A Serious Game for children with Autism Spectrum Disorder as a tool for play therapy. *2017 IEEE 5th*

International Conference on Serious Games and Applications for Health (SeGAH).

doi: 10.1109/SeGAH.2017.7939266

- Rubin, K. H. (2001). *The play observation scale (POS)*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.657.7522&rep=rep1&type=pdf>
- Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened Test Anxiety Among Young Children: Elementary School Students' Anxious Responses to High-Stakes Testing. *Psychology in the Schools, 50*(5), 489-499.
- Toplak, M. E., Bucciarelli, S. M., Jain, U., & Tannock, R. (2008). Executive functions: performance-based measures and the behavior rating inventory of executive function (BRIEF) in adolescents with attention deficit/hyperactivity disorder (ADHD). *Child Neuropsychology, 15*(1), 53-72.
- Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., . . . Pandey, A.K. (2019). Second Language Tutoring using Social Robots: A Large-Scale Study. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. doi:10.1109/HRI.2019.8673077
- von der Embse, N., Jester, D., Roy, D., & Post, J. (2017). Test anxiety effects, predictors, and correlates: A 30-year meta-analytic review. *Journal of affective disorders, 227*, 483-493.
- Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology, 7*(2), 131-149.
- Werkgroep en Steungroep Kleuteronderwijs. (2013). *Zwartboek Kleuters in de knel!* Retrieved from <https://www.wsk-kleuteronderwijs.nl/wp-content/uploads/2013/04/Zwartboek-Kleuters-in-de-knel-april-2013-LR.pdf>



Chapter 7

Conclusion and valorisation

7.1 Conclusion

The importance of early childhood development has long been established in research, which inspired the development and implementation of cognitive interventions. It has been clear that early childhood forms the foundation for future learning and school success, meaning children who are at risk of developmental delays should be stimulated as much as possible. In order to fully understand the cognitive abilities of children and how to stimulate these, the importance of testing and assessment techniques is needed in order to effectively, reliably and validly measure (changes in) the cognitive abilities. These tests are crucial in showing the current skill level in children, in tracking skill development (in relation to stimulation/interventions) and in making peer comparisons (i.e. establishing a norm). However, testing poses a particular problem for early childhood since measurements in this age group are difficult to operationalize. Usually, standardized conditions and tests are used to measure cognitive skills. In kindergarten, right before children start formal primary education, testing is of high importance but kindergarteners are not yet able to sustain a full standardized testing session. Specifically, certain cognitive skills, namely executive functions, needed to be able to self-regulate behaviour, are still being developed by kindergarteners.

This thesis focused on understanding cognitive abilities in kindergarteners by studying the effectiveness of cognitive intervention and studying innovative use of toys as an assessment tool. Regarding the latter, given that we live in a world where data are continuously collected and where technological applications can pose alternative and feasible solutions, this thesis explored whether it is possible to devise assessments that are more reliant on play and natural behaviour via the use of technological applications. Since play is the steering mechanism for child development, in this thesis, technological advances concerning toys are utilized to overcome strenuous testing situations in young children. This rationale is related to the concept of stealth assessment, in which learner performance data are gathered during learning in a playful way, such as a computer game. Therefore, within this concept of stealth assessment, technology plays a significant role as well (Shute, Ventura, Bauer & Zapata-Rivera, 2009). A focus on more natural, playful behaviour of young children is utilized, which is more in line with what educators and researchers agree to be the driving force of child development. On the spectrum from traditional test to free play, we wish to measure closer to free play, whilst maintaining objective and precise measurements. This has implications as to which cognitive

skills can be (no longer) measured. In this thesis, the different studies take these steps to answer the primal research question.

The study in Chapter 2 was aimed at understanding the current situation and effectiveness in educational practice regarding cognitive ability and intervention. It underlines the importance of assessment practices. Cognitive abilities of kindergarteners are, in practice, mostly understood as performance on standardized school tests, related to math and language. The study described in Chapter 2 demonstrated the relation between preschool education provided to young children before entering kindergarten and their scores on standardized school tests at the end of kindergarten at 5-6 years old. The results indicated a positive relationship between the intervention and test scores of the children. When looking specifically at target-group children (i.e. children of low educated parents), the relationship appears to be stronger, but relations are more noisy because the sample size becomes smaller.

In Chapter 3 we utilized a well-known framework in developmental psychology to understand specific cognitive abilities, namely executive functions (EF). We linked children's performance on traditional, existing EF tasks to their ability to solve mazes by measuring how they manoeuvred a small, programmable robot, the Bee-Bot, through it. The results showed that how children program the Bee-Bot, is related to problem-solving abilities, memory, visuospatial abilities and attention. With the exception of verbal fluency skills, these results were in agreement with studies that examined whether a robotics or programming intervention leads to an improvement of these specific skills. This means that while children play with the Bee-Bot and solve assignments, insights can be gained into these specific EF. This information is highly valuable for teachers who (would) like to use the Bee-Bot in their kindergarten classrooms, as targeted use of the tool can provide insights into these skills.

We continued this rationale in Chapter 4, where we designed a new toy that is capable of collecting data itself, Clever Maze (CM). CM looks like a regular maze but is, in fact, capable of collecting data on the behaviour of children while they solve mazes. Children move a magnetic ball through the maze via the use of a pen on top of a glass-looking surface. CM collects info on the coordinates and time of the ball inside the maze pattern. In this chapter, we conducted a proof of concept with a small number of children. Measurement accuracy, design and reactions of the participants are discussed in this chapter and are deemed to be very satisfactory.

In chapter 5 we studied whether performance of kindergarteners on existing EF tests is related to solving mazes in the CM. We extracted variables from the CM on how the children solved the mazes, which we then related to EF. The chapter discusses the findings with respect to convergent and discriminant validity of assignments using CM. The relationship with performance on this CM seems to be strongest with visuospatial reasoning and visuospatial memory, problem-solving and to some extent working memory.

In the study in chapter 6, we let children play with CM, but without giving them explicit instructions to solve mazes. The aim of this study was to move even further away from the traditional viewpoint around early childhood testing and to explore whether “testing without testing” is at all possible. By just providing the toy to children and not providing instruction, we expected normal play behaviour to unfold. We compared the behaviour of children with the CM in this free play situation to their performance on standardized EF tests and parent’s responses on a questionnaire pertaining to their child’s everyday behaviours related to EF. The play of the children in CM related significantly to planning abilities, problem-solving and working memory skills as measured by EF tasks. We also see a relationship with everyday behaviour as measured by the BRIEF in terms of initiative-taking and flexibility.

We can conclude that it is possible to find relations between how children handle programming a robot and playing with a smart toy to specific EF in the kindergarten age group. Innovative means can provide rich data that are collected in a more natural fashion, thus decreasing potential test-related stress, as well as time and expertise investments from the examiner. More studies using CM and other new tools could investigate whether a comprehensive view on early childhood EF and other skills is possible, perhaps even information on language and math skills. Higher diversity of smart toys is needed as well as repeated studies in different age groups in order to receive a comprehensive assessment of cognitive skills. Eventually, when sufficient information has been collected, an expansion to more clinical populations could be made. The biggest established advantage of smart toys is the richness of the collected data. This poses ample opportunities for research, as detailed information on processes while playing or solving puzzles provides valuable insights beyond existing fail/pass criteria.

7.2 Implications and Valorisation

Chapter 2 provides valuable insights for policymakers and educators alike regarding early childhood interventions, as well as creates an understanding of the use of academic achievement tests in kindergarten. Chapters 3, 4, 5 and 6 all focus on assessment in young children in relation to recent technological advances. Whilst Chapter 3 provides insight for educators directly on materials available for purchase, Chapters 4-6 answer more fundamental research questions on a newly developed tool.

7.2.1 The Relation Between Preschool Education and Cognitive Outcomes on Standardized School Tests

The results of Chapter 2 are relevant for policymakers regarding the investments in preschool cognitive interventions. In the Netherlands, there has been ongoing debate about the effectiveness of the VVE program as a whole (Fukkink, Jilink & Oostdam, 2017). The current study adds to this debate as a unique dataset, the OML ('Onderwijsmonitor Limburg'), was used for the analysis. The results yield the strongest relationships between an educational intervention for the children who are part of the target group based on parental education level. They also highlight the difficulty with which target-group children can be reached in survey research. Furthermore, this study shows there was a positive trend for all children receiving additional support in comparison to children who did not receive PEP, even if they were not part of the target group. Thus, findings are relevant in understanding the added-value of educational intervention, not just for target-group children. This is closely related to one of the four dilemmas mentioned by Coppens (2015) about investments made in the Netherlands. If the aim is to close the gap, the intervention should be targeted specifically at children from disadvantaged households. If the aim is to provide richer opportunities, more if not all children should get access to the intervention. Therefore, this study also highlights why it is important to understand cognitive abilities in young children.

7.2.2 The Relation Between Executive Functioning Skills and Programming a Bee-Bot

The study in Chapter 3 poses a clear contribution to the scientific literature on the potential of using programmable toys or robots for young children as an assessment tool. Most applications of similar tools are oriented toward cognitive stimulation, whereas this study takes a different perspective and focuses on assessment. Furthermore, the results are very informative for

schools and educators who utilize or are interested in utilizing Bee-Bots and similar tools, because it is a very approachable tool. The Bee-Bot is a very easy-to-use tool specifically designed for young children and can be purchased by schools themselves. Not only is this type of material appealing and interesting for young children, expanding the potential of its application and increasing the understanding of underlying cognitive concepts is highly valuable. The set-up of the study appealed to many teachers who participated in the study. If they were already in the possession of the Bee-Bot or similar materials, they reported to the researcher that they were interested to know what the materials could actually be used for other than fun exercises with the children. Schools often purchase these materials in line with the increased focus on STEM education (Bybee, 2010), where it is becoming increasingly established that familiarity with programming and robotics concepts from an early age onward is beneficial for skill development. However, deepened knowledge of the potential of these tools has yet to be acquired (Bers, Seddighin, & Sullivan, 2013). Studies such as the current one are steps in this direction and have much support from the educational field.

7.2.3 Development of a Smart Toy as a Measurement of Executive Functioning Skills

Before performing a wider data collection in kindergarteners with a new assessment tool, it is important to perform pre-testing. We pre-tested a Clever Maze (CM). This pre-test allowed for checking whether the material is suitable and safe for children, whether the difficulty level is accurate and the material has appeal to the children. It also provided an opportunity to compare the tool's accuracy as compared to a camera-based approach, thus providing a proof of concept that follow-up studies can be performed.

7.2.4 Children's Executive Functioning: What Can Solving Mazes and Free Play on Clever Maze Reveal?

The two studies described in Chapter 5 and 6 investigated the link between children's performance on traditional EF measures and Clever Maze (CM) assignment and free play outcomes. Taking the findings of Chapter 5 and 6 together, there is an overlap in the EF that are picked up by CM when providing an assignment or when letting children play freely. Specifically problem-solving and working memory skills are used in both situations. Regarding the assessment of cognitive abilities in young children, there is a need for innovative testing materials. Technology provides us with many new possibilities in this regard, as a tool that is smart enough to collect data on its own saves on time and expertise investments. Not

underestimating the value of screen-based assessments or pencil-and-paper assessments, smart toys pose the possibility to collect data on behaviour in real-life, potentially without any hassle of observation. Smart algorithms could be used on the collected data (automatically), providing direct insights into child EF.

In more traditional neuropsychological tests, information on subprocesses of solving tasks are usually discarded to receive an eventual score or outcome. This score is compared to a norm score provided by the test developers. In the current studies, we researched a sample of children developing within the normative range and studied extracted information on the processes of solving the assignments or playing with CM. This line of thinking is highly valuable as it underlines more process-related understanding of abilities and development, rather than an absolute thinking in terms of pass and fail. For instance, becoming faster at solving assignments is also a sign of growth, which is usually not incorporated into pass/fail items.

The idea of testing without testing poses the most innovative question of the thesis. In a world where even fridges are becoming 'smart' and information is continuously collected on human behaviour, technology could be implemented in toys to assess behaviour and skills of children. In the past, the connection between toys and technology has often been pursued to create more interaction between toy and child for commercial purposes, such as Hello Barbie (Mattel), or to enhance cognitive abilities with respect to an intervention, such as L2TOR for second language learning (Belpaeme et al., 2018). The continuous collection of data on children, however, can also provide valuable insights into current skill levels and eventually in developmental trajectories. If sufficient toys in a classroom are smart, such as puzzles, mazes, dolls, blocks and so on, a comprehensive view on EF and other skills may be possible. The central idea is that natural behaviour of the child can unfold without an examiner testing the child in a one-to-one setting and specific tasks having to be completed.

If a tool like CM is present in a classroom, for instance, it is available for repeated use and different maze patterns could be used. The potential combination of this tool with other smart toys can provide a comprehensive insight into child cognitive ability. Developmental patterns could be easily plotted and understood on an individual level. Such 'wild' ideas can initially only take place in the lab of research institutions or controlled data collections in schools with a researcher present and are currently too expensive to create en masse. However,

if well-established by research and repeated studies in multiple age groups, commercial applications of such products in collaboration with researchers and educators could yield highly valuable insights and innovations with low investments.

7.3 References

- Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., . . . Oudgenoeg-Paz, O. (2018). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics, 10*(3), 325-341.
- Bers, M., Seddighin, S., & Sullivan, A. (2013). Ready for robotics: Bringing together the T and E of STEM in early childhood teacher education. *Journal of Technology and Teacher Education, 21*(3), 355-377.
- Bybee, R. W. (2010). Advancing STEM education: A 2020 vision. *Technology and Engineering Teacher, 70*(1), 30.
- Coppens, K. M. (2015). *Het effect van Voor- en Vroegschoolse Educatie in vier dilemma's*. Retrieved from <https://www.educatieveagendalimburg.nl/bijdragen/schoolkeuzes/het-effect-van-voor-en-vroegschoolse-educatie-vier-dilemmas>
- Fukkink, R., Jilink, L., & Oostdam, R. (2017). A meta-analysis of the impact of early childhood interventions on the development of children in the Netherlands: an inconvenient truth? *European Early Childhood Education Research Journal, 1-11*.
- Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah, NJ: Routledge, Taylor and Francis.



Appendices

Appendix A

Table A. Extended overview of multiple regressions analysis in table 2.4 (N = 4075) to investigate how well preschool education predicts test outcomes; (1) without control variables, (2) with gender, age in months, school year, municipality, parental education level, parental country of birth and language spoken at home, whether the child attended a daycare facility, and which school the child attended as predictor variables and (3) replacing school with municipality as a predictor variable.

OLS results of all children in the sample						
	Language (1)	Math (1)	Language (2)	Math (2)	Language (3)	Math (3)
PEP	0.67** (0.33)	0.86** (0.36)	0.65** (0.33)	1.08*** (0.36)	0.54* (0.33)	0.77** (0.36)
Gender			1.89*** (0.30)	-0.45 (0.33)	1.81*** (0.30)	-0.49 (0.34)
Age			0.35*** (0.04)	0.46*** (0.04)	0.31*** (0.04)	0.43*** (0.04)
Daycare use			0.56 (0.56)	0.46 (0.62)	0.047 (0.57)	-0.02 (0.63)
Isced ^a						
2			0.82 (1.85)	3 (2.07)	0.67 (1.87)	2.48 (2.08)
3			2.49 (1.79)	5.49*** (2)	2.53 (1.81)	5.34*** (2.02)
4			5.46*** (1.79)	7.91*** (2)	5.66*** (1.81)	7.97*** (2.02)
5			6.77*** (1.81)	9.45*** (2.02)	6.70*** (1.83)	9.39*** (2.04)
Parental country of birth ^a						
2			-0.52 (0.60)	-0.48 (0.67)	-0.69 (0.60)	-0.73 (0.67)
3			-4.21*** (1.16)	-5.35*** (1.29)	-3.71*** (1.16)	-4.70*** (1.30)
Language at home ^a						
2			-1.87** (0.81)	-1.22 (0.90)	-2.33*** (0.82)	-1.93** (0.91)
3			-5.90*** (2.01)	0.45 (2.24)	-6.56*** (2.02)	-0.66 (2.26)
School year ^a						
2			1.03** (0.45)	0.57 (0.50)	1.16*** (0.45)	0.73 (0.50)
3			0.59 (0.44)	0.69 (0.49)	0.80* (0.43)	0.90* (0.48)
4			0.57 (0.43)	0.42 (0.48)	0.68 (0.42)	0.70 (0.47)
Constant	68.21*** (0.20)	87.44*** (0.22)	37.62*** (3.28)	49.03*** (3.67)	42.47*** (3.04)	50.89*** (3.79)
R Squared	0.00	0.00	0.17	0.16	0.09	0.09

Note. Standard errors in parentheses. * p < .1, ** p < .05, *** p < .01. In model (2), school level was added as a predictor variable, containing 120 schools. In model (3), municipality was added with 17 levels. To avoid abundant tables, model (2) and (3) were therefore cropped. The full table can be requested with the corresponding author. a. The first category of the variable serves as a baseline for analysis. Target group children are defined according to parental education level criteria.

Table B1. Extended overview of multiple regressions analysis in table 2.5.1 to investigate how well preschool education predicts language test outcomes; (4) without any added predictor variables comparing target group and non-target group children according to parental education level criteria, and (5) using all control variables as in analysis model (3) whilst making the same comparison between target group and non-target group children.

OLS results of target (TG) and non-target (non-TG) group children for language				
	Language TG (4a)	Language Non-TG (4b)	Language TG (5a)	Language non- TG (5b)
PEP	3.34* (1.85)	0.56* (0.33)	1.46 (1.94)	0.44 (0.34)
Gender			1.64 (1.67)	1.74*** (0.31)
Age			0.50** (0.19)	0.25*** (0.04)
Daycare use			3.10 (2.57)	0.40 (0.59)
Parental country of birth ^a				
2			7.91** (3.34)	-1.18* (0.63)
3			1.99 (3.64)	-4.89*** (1.26)
Language at home ^a				
2			-7.73** (3.56)	-1.71** (0.86)
3			-9.68 (6.32)	-5.52** (2.20)
School year ^a				
2			-0.30 (2.67)	1.15** (0.46)
3			1.75 (2.65)	0.89** (0.45)
4			-2.28 (2.61)	0.86** (0.44)
Constant	63.26*** (1.03)	68.39*** (0.20)	32.75** (15.08)	50.31*** (2.90)
N	128	3947	128	3947
R Squared	0.03	0.00	0.36	0.06
All control variables	no	no	yes	yes

Note. Standard errors in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. ^aThe first category of the variable serves as a baseline for analysis. Target group children are defined according to parental education level criteria.

Table B2. Extended overview of multiple regressions analysis in table 2.5.2 to investigate how well preschool education predicts math test outcomes; (6) without any added predictor variables comparing target group and non-target group children according to parental education level criteria, and (7) using all control variables as in analysis model (3) whilst making the same comparison between target group and non-target group children.

OLS results of target (TG) and non-target (non-TG) group children for math				
	Math TG (6a)	Math non-TG (6b)	Math TG (7a)	Math non-TG (7b)
PEP	3.46* (1.89)	0.75** (0.37)	0.85 (2.20)	0.68* (0.37)
Gender			-1.27 (1.89)	-0.57 (0.35)
Age			0.48** (0.21)	0.38*** (0.04)
Daycare use			1.39 (2.91)	0.36 (0.66)
Parental country of birth ^a				
2			0.04 (3.78)	-0.96 (0.70)
3			-1.91 (4.11)	-5.94*** (1.40)
Language at home ^a				
2			-3.53 (4.02)	-1.53 (0.95)
3			-1.24 (7.14)	0.27 (2.45)
School year ^a				
2			-0.22 (3.02)	0.64 (0.51)
3			-1.60 (2.99)	1.02** (0.50)
4			-1.78 (2.95)	0.81* (0.49)
Constant	81.97*** (1.06)	87.63*** (0.22)	49.60*** (17.00)	61.00*** (3.24)
N	128	3947	128	3947
R Squared	0.03	0.00	0.22	0.05
All control variables	no	no	yes	yes

Note. Standard errors in parentheses. * p < .1, ** p < .05, *** p < .01. ^aThe first category of the variable serves as a baseline for analysis. Target group children are defined according to parental education level criteria.

Appendix B

Handleiding en scoring voor de Bee-Bot taken

Materiaal:

- Doolhof mat
- Bee-Bot (plus oplader)
- Scoreformulier
- Pen
- 2 kussentjes
- Stopwatch
- Doek om doolhof af te dekken

Opzet:

Kind wordt gevraagd om op het kussentje op de grond te gaan zitten dat vóór de doolhof ligt en geïnstrueerd hier te blijven zitten gedurende de testafname. Om te voorkomen dat het kind afgeleid wordt door dingen die op de gang gebeuren, zullen het kussen en doolhof zodanig worden neergelegd dat het kind met de rug naar de deur/raam gericht zit. De onderzoeker zit op een kussentje naast het kind gedurende de afname. De verbale instructies zijn schuingedrukt, handelingen staan aangegeven tussen sterretjes.

Instructie: *Je mag daar (*wijst naar kussen*) op het kussentje gaan zitten.*

Indien het kind gedurende de afname opstaat: *Zou je weer terug op het kussentje willen gaan zitten?*

Afname:

Het is de bedoeling dat het kind de Bee-Bot zodanig programmeert dat hij van de ingang van de doolhof naar de uitgang gestuurd wordt. De sessie bestaat uit drie opdrachten met ieder een verschillende doolhof/design die toenemen in moeilijkheid (Zie afbeeldingen). Zoals op de afbeeldingen te zien is, zijn er doodlopende gangen die dienen als afleiders. Het doolhofdesign voor de eerste opdracht zal bij binnenkomst al klaarstaan (maar afgedekt met een doek), de designs voor de volgende twee opdrachten zullen tijdens de afname opgebouwd worden. Het kind krijgt per poging 2 minuten de tijd vanaf het moment dat de instructie is afgerond.

Instructie:

Een paar weken terug hebben we met de Beebot gespeeld op de Lettermat, weet je nog? Vandaag gaan we weer met de Bee-Bot spelen, maar nu met een doolhof. (*Doek van doolhof halen*).

Ik zal Bee-Bot al bij het begin neerzetten. (*Bee-Bot oppakken en neerzetten bij het 2^o hokje van de onderste rij*). *Nu moet jij kijken waar de uitgang is en welke stapjes Bee-Bot moet zetten om daar te komen. Zie je de uitgang? Wijs eens aan?*

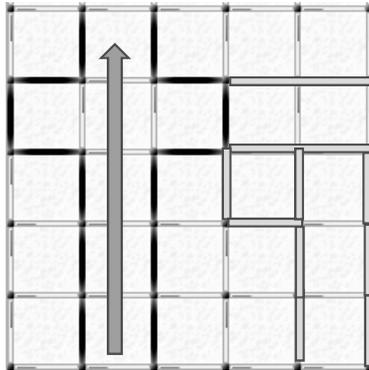
Het is de bedoeling dat jij de Bee-Bot de goede instructies geeft, je mag Bee-Bot dus niet

onderweg nog opnieuw programmeren. Doe goed je best. Het is niet erg als je een fout maakt. Is er nog iets dat je niet begrijpt?

Doolhof 1

Instructie: *Klaar? Begin maar! (*Start de tijd*)*.

Als het kind klaar is: *Goed gewerkt, nu gaan we de volgende proberen.*



1. Rechtdoor.

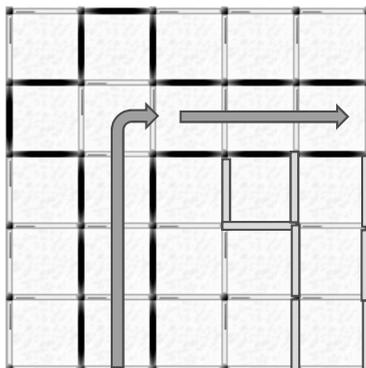
Doolhof 2

De 2^e doolhof zal opgebouwd worden waar het kind bij is.

Instructie: *Hier zie je de volgende doolhof, nu mag je weer hetzelfde doen als bij de eerste. Dus we beginnen weer bij de ingang (*wijst naar begin*) en jij mag Bee-Bot naar de uitgang brengen. Kun je de uitgang aanwijzen?*

*Klaar? Begin maar. (*Start de tijd*)*

Als het kind klaar is: *Goed gewerkt, nu gaan we de volgende proberen..*



2. Bocht 90 graden.

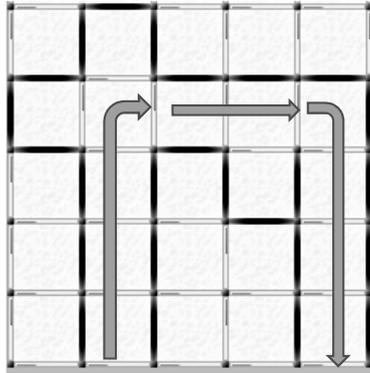
Doolhof 3

De 3^e doolhof wordt opgebouwd.

Instructie: *Hier zie je het laatste doolhof. We beginnen weer bij de ingang en jij mag Bee-Bot naar de uitgang brengen. Kun je de uitgang aanwijzen?*

*Klaar? Begin maar. (*Start de tijd*)*

Als het kind klaar is: *Goed gewerkt, dit was de laatste opdracht met de Bee-Bot.*



3. 2 bochten 90 graden.

B

Mogelijke problemen bij start:

- Mocht het kind niet van zelf beginnen na 30 seconden, zeg: *Met welke stappen moeten we beginnen?* (Tenzij het kind de indruk geeft dat het nog aan het nadenken is).
- Mocht het kind niet weten wat te doen: *Probeer het maar eens.*

Mogelijke problemen tijdens de opdracht:

- Als het kind de Bee-Bot foutief heeft geprogrammeerd, zeg: *Oei je bent niet bij de uitgang geraakt, maar je mag het nog eens opnieuw proberen.* (*Bee-Bot weer op beginplek zetten en van voor af aan beginnen*)
- Kind zegt dat het hem/haar niet lukt, zeg: *Probeer het maar.*
- Kind vraagt hulp, zeg: *Ik wil zien hoe je het alleen kunt.*
- Wanneer het kind 2 keer vergeet om geheugen te resetten, zeg: *Eerst even het hoofdje leegmaken.*
- Wanneer het na 2 keer proberen nog niet lukt; doorgaan naar volgende doolhof. Zeg: *Goed geprobeerd, nu gaan we het met een andere opdracht proberen.*
- Afbreekregel: Wanneer het kind 2 opeenvolgende doolhoven niet opgelost krijgt: stoppen met opdrachten. Zeg: *Goed gewerkt, we zijn klaar met het doolhof.*

Scoring:

De volgende variabelen zullen genoteerd worden op het scoringsformulier tijdens de afname:

- Denktijd: Deze zal bijgehouden worden met een stopwatch. De tijd start vanaf het startsignaal en eindigt wanneer het kind de eerste knop indrukt.
- Uitvoeringstijd: Deze zal bijgehouden worden met een stopwatch. De tijd start vanaf het

moment dat het kind de eerste knop indrukt totdat de Bee-Bot zijn eindbestemming bereikt heeft (2 piepjes).

- Succesvolle uitvoering: Noteer ja/nee of het kind de Bee-Bot dusdanig geprogrammeerd heeft dat hij op de goede eindbestemming terecht is gekomen.

- Fouten: Noteer aantal door middel van turven. Opgedeeld in:

* Knop teveel ingedrukt.

* Knop te weinig ingedrukt.

* Foutieve stap: Foutieve knop ingedrukt.

* Clear button vergeten.

* Draaiknop verkeerd gebruiken: gebruikt draaiknop als stap naar links/rechts.

* Denkt tussendoor na over de volgende stap: Stuur Bee-Bot eerst naar een gedeelte van de doolhof en programmeert dan de rest.

- Fysieke gedrag: Verschillende gedragsmatige handelingen die uitgevoerd zijn zullen worden aangevinkt op het scoringsformulier.

<u>Tijd</u>	Opdracht 1	Opdracht 2	Opdracht 3	<u>Succesvolle uitvoering</u>
Denktijd				Opdracht 1: Ja / Nee
Uitvoeringstijd				Opdracht 2: Ja / Nee
Totale tijd				Opdracht 3: Ja / Nee

<u>Fouten</u>	Opdracht 1	Opdracht 2	Opdracht 3
Knop teveel ingedrukt			
Knop te weinig ingedrukt			
Foutieve knop ingedrukt			
Draaiknop gebruikt als stap naar links/rechts			
Verdeelt/programmeert de opdracht in subdelen			

<u>Gedragmatige handelingen</u>	Opdracht 1	Opdracht 2	Opdracht 3
Bewegen van lichaam (handen etc).			
Stappen tellen die gezet moeten worden.			
Hardop nadenken.			
Voert de stappen al in tijdens het tellen/nadenken.			
Kan alle stappen achter elkaar programmeren zonder lange denkpauzes tussendoor (20 seconden tussen stappen).			
Kind heeft tijdens het programmeren zelf door dat het fout is en begint opnieuw met programmeren (clear button).			
Oriëntatie ten opzichte van de doolhof verandert.			
Kind vraagt bevestiging of hij/zij het goed doet.			
Afgeleid tijdens het uitvoeren van de taak.			
Bee Bot oppakken en zo de stappen tellen die gezet moeten worden.			

Overige handelingen/opmerkingen

Appendix C

Handleiding Clever Maze testafname: Vrij spel en opdrachten

1. Algemeen:

In de koffer zitten de volgende benodigdheden:

- 1 doolhof (genummerd 1 of 2)
- 3 doolhofborden (genummerd 1, 2 en 3)
- 1 accu
- 1 lader voor de accu
- 1 magnetische pen (en eventueel reserve)
- Eventueel reserve magnetische balletjes
- Een afstandsbediening met usb-aansluiting of tablet

Set-up:

Er wordt een tafeltje op kinderhoogte voorzien (vraag dit aan de juf indien het niet in het lokaal beschikbaar is). Hierop wordt Clever Maze geplaatst. De ruimte wordt zoveel mogelijk vrij gemaakt, geen ander speelgoed of afleiding in de buurt. De laptop wordt op een afstand van het speelgoed gezet, met het scherm naar de onderzoeker gericht. De onderzoeker zit aan het tafeltje, terwijl het kind ervoor staat. De verbale instructies staan schuingedrukt, handelingen staan aangegeven tussen sterretjes.

2. Vrij spel meting:

- a) *Zeg: Jij mag hier voor het tafeltje gaan staan. Ik heb een aantal spelletjes voor jou bij en hier gaan we mee beginnen. (*Positioneer het kind voor het doolhof*). We beginnen met een doolhof, weet je wat dat is? (de meesten kennen het wel, als het kind het niet kent, herkent het het normaal wel wanneer je het doolhof onthult).*
- b) **Haal het andere bord dat je als afdekplaat gebruikte van het doolhof af*.*
- c) **Start de afname door op “start” te klikken op de afstandsbediening, dus nog voor je het kind de pen geeft*.*
- d) *De 2^e led-lampje van rechts gaat dan van groen naar oranje.*
- e) *Zeg: Met deze pen kun je het balletje in dat doolhof bewegen. En je mag gewoon spelen.*
- f) **Geef het kind de pen en start de stopwatch.**

- g) Het kind mag maximum 5 minuten spelen (Zie hieronder voor aanpak verschillende situaties).
- h) *Na afloop van de exploratie druk je op ‘stop’ en zet je ook de stopwatch stil. De tijd noteer je enkel wanneer het kind geen 5 minuten lang kon spelen.*
- i) *Zeg: Kun je me vertellen wat je precies hebt gedaan tijdens je spel?* *Noteer het antwoord op het scoreformulier, zelfs als het kind niets zegt of als het kind zegt dat het het niet weet.*
- j) Je laat het exploratie bord in het doolhof zitten.

Belangrijk: Kind aanmoedigen om het balletje te bewegen, zonder het een doel te geven. Dus zeg niet: ga van daar naar daar. Je noteert je observaties; zo let je bijvoorbeeld op: praat het kind tijdens het spelen, speelt het kind erg rommelig of net heel netjes, zucht het of lacht het, maakt hij/zij er een verhaaltje van, zie je expliciet dat het kind probeert op een bepaalde plek te geraken, etc. Als het kind vraagt wat de bedoeling is of wat het moet doen, herhaal je dat het gewoon mag spelen.

Mogelijke problemen bij start:

- Als het kind niets doet moedig je het na 30 seconden aan om iets te doen.
- Als het het kind echt niet lukt om met de pen het balletje te bewegen of als het kind iets anders doet, doe je het na 90 seconden voor.

De bedoeling is dat je dit enkel doet als je ziet dat het kind echt vastloopt. Sommige kinderen zullen uit zichzelf vragen wat de bedoeling is, of waar ze heen moeten. Dan is het net belangrijk dat je het kind niet stuurt.

Mogelijke problemen tijdens de afname:

- Het kind beweegt zo ruw of speels met de pen over het doolhof dat het contact met de bal niet geregistreerd wordt. Sommige kinderen proberen dit even. Pas wanneer het gedrag aanhoudt, zeg je: *Probeer met de pen goed op het balletje te blijven.*
- Het kind laat het balletje bewegen terwijl de pen op een afstandje is van de bal (horizontale of verticale afstand), waardoor er niets geregistreerd wordt. Opnieuw geldt dat je pas ingrijpt als dit gedrag even aanhoudt, je zegt: *Probeer met de pen goed op het balletje te blijven.*
- Het kind legt een hand op het doolhof terwijl het speelt. Voor de meting zou dit niet direct een gevaar zijn, kijk vooral goed naar de led-lamp die aangeeft of de bal geregistreerd wordt. Indien die heel sterk knippert zeg je tegen het kind: *Haal je andere hand maar van het doolhof af, dan kun je beter spelen.*

- Het kind steunt met de ellebogen op het doolhof: hier grijp je onmiddellijk in en vraag je of het kind de ellebogen eraf wilt halen.

Problemen bij stop:

- Indien het kind langer speelt dan 5 minuten, zeg je na 5 minuten dat het tijd is om wat anders te gaan doen.
 - o Instructie:

Ok goed gespeeld, je mag de pen teruggeven aan mij. En dan vraag je wat het kind precies deed tijdens het spel om te noteren op het scoreformulier, zoals aangegeven bij 2.i.

- Indien het kind niet langer dan 2 minuten gespeeld heeft, zeg je dat het nog even tijd heeft als het nog wat zou willen spelen.
- Indien het kind stopt met spelen tussen 2 en 5 minuten zeg je zelf nog niets en wacht je tot het kind verbaal of op een andere manier duidelijk aangeeft dat het klaar is.

3. Opdrachtenmeting:

Dek het doolhof af met de twee afdekplaten (kwart en half), ook al ligt het exploratiebord eronder. Je zorgt dat de rode en gele ster zichtbaar zijn (kwadrant 1 vanuit perspectief kind rechts boven). Zeg: *Nu ga ik je vragen om iets te doen met het doolhof, dus goed opletten.*

Verwissel dan pas de borden, stop het opdrachtenbord in het doolhof met het balletje in het midden.

4	1
3	2

Vanuit perspectief van het kind ziet het er als volgt uit, met kwadrant 1 de makkelijkste, 2 gemiddeld, 3 een afleidende hoek en 4 de moeilijkste opdracht. Initieel is dus enkel 1 zichtbaar. 2 is afgedekt met een kwart-afdekplaat, 3 en 4 met een halve-afdekplaat (de afdekplaten zijn dunne plankjes die op maat van het doolhof gesneden zijn, eentje bedekt een kwart van het bord; een andere de helft).

Start vervolgens de meting.

- a) Zeg: *De bal ligt nu bij de rode ster (*wijs aan*), kun je hem naar de gele ster (*wijs aan*) brengen?.* *Terwijl je dit zegt druk je op marker 1*

- b) Wanneer het kind dit gedaan heeft, zeg je: *Kun je de bal terugbrengen van de rode ster (*wijs aan*) naar de gele ster (*wijs aan*)?* *Tegelijkertijd druk je op marker 2*
- c) *Dan haal je de kwart-plaat weg die kwadrant 2 bedekt, zodat de groene ster ook zichtbaar wordt*
- d) Zeg: *Kun je de bal nu van de rode ster (*wijs aan*) naar de groene ster (*wijs aan*) brengen?* *tegelijkertijd druk je op marker 1*
- e) Bij succesvolle uitvoering zeg je: *Kun je de bal terug van de groene ster (*wijs aan*) naar de rode ster (*wijs aan*) brengen?* *Tegelijkertijd druk je op marker 2*
- f) *Vervolgens haal je de halve-plaat weg, zodat de blauwe ster en perseveratieve cirkel zichtbaar worden*
- g) Zeg: *Kun je de bal nu eens van de rode ster (*wijs aan*) naar de blauwe ster (*wijs aan*) brengen?* *druk terwijl je dit zegt op marker 1*
- h) Bij succesvolle uitvoering, zeg je: *Kun je de bal eens terugbrengen van de blauwe ster (*wijs aan*) naar de rode ster (*wijs aan*)?* *druk tegelijkertijd op marker 2*
- i) *Als het kind dit gedaan heeft, druk je op 'stop'*

Noteer interessante observaties, zoals verbale uitlatingen, kinderen die het erg lastig hebben met de opdracht of het net heel erg vlot oplossen.

Problemen bij opdrachten:

- Als de eerste opdracht niet lukt, probeer je de tweede nog. Als deze ook niet lukt, stop je de afname. Als de tweede wel zou lukken, neem je ook de derde af. Je geeft geen expliciete feedback aan de kinderen of ze het goed of slecht hebben gedaan.
- Tijdens deze afname wordt de tijd niet met de stopwatch gevolgd. Moest een opdracht niet lukken is er geen tijdslimiet, het kind mag blijven proberen tot het zelf opgeeft.
- Als een kind heel snel opgeeft, mag je het aanmoedigen om nog eens te proberen. Noteer dit dan zeker op het scoreformulier.



Nederlandse samenvatting
(Dutch summary)

Motivatie en doel

Onderzoek heeft aangetoond dat de vroege kindertijd een belangrijke periode is voor de cognitieve ontwikkeling. Dit komt met name door neurologische flexibiliteit, zowel structureel als functioneel, waardoor jonge kinderen gevoelig zijn voor invloeden van buitenaf (Brown & Jernigan, 2012). Door deze onderzoeksbevindingen is men meer gaan nadenken over hoe de ontwikkeling van jonge kinderen gestimuleerd kan worden door middel van interventies. Dergelijke interventies kunnen met name betekenisvol zijn voor kinderen met een ontwikkelingsachterstand. Maar hoe weten we nu precies hoe het staat met de cognitieve vaardigheden van een kind? Een bijkomend vraagstuk is dan ook welke testen gebruikt kunnen worden om zich te krijgen op de vaardigheden van jonge kinderen. Ontwikkelde testen dienen betrouwbaar en valide te zijn om zo de beoogde vaardigheden te meten. Deze testen vragen daarom veelal gestandaardiseerde condities.

Inzicht in de ontwikkeling van vaardigheden van kinderen is een belangrijk onderwerp in de kleuterperiode, wanneer kinderen naar school beginnen gaan. Binnen het onderwijs wordt echter spanning ervaren tussen de manier waarop kinderen leren en de manier waarop getoetst wordt. Het is algemeen bekend dat kinderen veel baat hebben bij spelen en dat spel een belangrijke catalysator voor leren is (Goldstein, 2012). Spel wordt daarom voortdurend ingezet in de vroege kindertijd als leer methode, zowel thuis als op school. Alleen al de inrichting van de klas, met een goed doordachte manier van aanbieden van speelgoed, nodigt kinderen uit tot leerrijk spel. Spel wordt soms ook al getoetst, bijvoorbeeld door middel van systematische observaties, of vragenlijsten voor leerkrachten en ouders. Dankzij technologische ontwikkelingen bestaan er echter veel meer mogelijkheden om informatie te verzamelen over het spel van kinderen. ‘Slim’ speelgoed wordt momenteel soms gebruikt om de cognitieve vaardigheden van kinderen te versterken, maar zou ook als test gebruikt kunnen worden.

Gegeven het belang van vroege meting en de ervaren spanning tussen de beleving van kleuters en gestandaardiseerde testen, wordt in deze thesis onderzocht of het mogelijk is om meetinstrumenten te ontwikkelen die meer tegen spel en natuurlijk gedrag aanleunen, met behulp van technologische toepassingen. In een eerste studie wordt de effectiviteit van een bestaande cognitieve interventie voor peuters met meer of minder kansen op een ontwikkelingsachterstand bestudeerd. Cognitieve vaardigheden wordt hier vooral begrepen

als de uitkomst op gestandaardiseerde toetsen op school, met een focus op taal en rekenen. In de studie werd onderzocht of een cognitieve interventie in de peutertijd samenhangt met hogere toetsscores in de kleutertijd. In de andere hoofdstukken worden innovatieve meetmethodes voor de cognitieve vaardigheden van kleuters onderzocht. Hierbij wordt het verband tussen gebruik van deze innovatieve meetmethodes en bestaande testen voor kleuters bekeken. De bestaande testen die gebruikt worden in de studies meten executieve functies (EF). EF zijn cognitieve vaardigheden die zorgen voor alledaagse gedragsregulatie. Uit onderzoek is gebleken dat ze erg belangrijk zijn voor verschillende uitkomsten in het leven en dus ook in het onderwijs, zoals schoolrijpheid (Blair & Razza, 2007) of reken- en leesvaardigheid (Borella, Carretti, & Pelegrina, 2010). Algemeen wordt aanvaard dat EF bestaan uit drie onderdelen, namelijk werkgeheugen, inhibitie en cognitieve controle (Diamond, 2013), hoewel bij kleuters die zich ontwikkelen vaak maar twee factoren gevonden worden (Viterbori, Usai, Traverso, & De Franchis, 2015). In de tweede studie werd het gebruik van een bestaand speelgoed onderzocht. Kinderen losten opdrachten op met een kleine, programmeerbare robot in de vorm van een bij. Veel basisscholen beschikken al over deze robot, vergelijkbare robots of spelletjes waarbij dezelfde principes gebruikt worden. Hoe kinderen de robot programmeerden werd vergeleken met hun uitkomsten op traditionele EF-taken.

Vervolgens werd als onderdeel van deze thesis een ‘slim’ speelgoed ontwikkeld dat in staat is om zelfstandig gegevens te verzamelen over het gedrag van kinderen. De Clever Maze (CM) ziet eruit als een gewoon houten doolhof. Door gebruik te maken van schermtechnologie worden gegevens verzameld over hoe kinderen doolhoven oplossen. In de derde studie werd de werking van de CM onderzocht met een klein aantal kleuters. In de vierde en vijfde studie werden vervolgens de CM-data van kleuters vergeleken met hun uitkomsten op traditionele EF-taken, waarbij in de vierde studie aan kleuters werd gevraagd om opdrachten te maken met de CM en ze in de vijfde studie geen opdracht kregen en vrij mochten spelen met de CM. In de vijfde studie werden de resultaten ook vergeleken met metingen op basis van een vragenlijst onder ouders over de alledaagse EF van hun kind.

Resultaten en implicaties

De eerste studie liet positieve verbanden zien tussen een cognitieve interventie en betere toetsresultaten voor alle kinderen die deelnamen aan de interventie. De winsten waren het

grootst voor doelgroepkinderen. Het is bemoedigend dat VVE het grootste effect lijkt te hebben voor de doelgroepkinderen. Daarnaast is het ook erg interessant dat zelfs voor kinderen die niet deel uitmaken van de doelgroep, er een positieve trend zichtbaar is. Dit resultaat is erg interessant voor onderwijsbeleid en mensen die actief zijn in het onderwijs zoals leerkrachten en schooldirecteuren. De tweede studie, waarbij gebruik gemaakt werd van een programmeerbare robot, toonde aan dat de programmeervaardigheden van kleuters samenhangt met bepaalde EF. Met name het non-verbaal redeneren, plannen, verbale vloeiendheid en geheugen hingen samen met de programmeervaardigheid van de kleuters. Gezien de robot die gebruikt werd in dit onderzoek vaak gekend is op scholen, is deze bevinding niet alleen waardevol voor onderzoekers. Door een verbeterd inzicht in welke vaardigheden geobserveerd kunnen worden wanneer kinderen omgaan met de aanwezige robots in de klas, kunnen dergelijke robots gerichter ingezet worden.

De derde studie toonde aan dat de CM effectief was in het verzamelen van accurate data over hoe kinderen een balletje door het doolhof laten bewegen. De vierde studie, waarin kinderen opdrachten maakten met de CM, toonde aan dat de CM waardevolle informatie verzamelt over EF. Met name visuospatieel geheugen en redeneren, probleemoplossend vermogen en werkgeheugen lijken gerelateerd aan hoe kinderen de CM-doolhoven oplossen. Wanneer kinderen geen opdracht kregen en vrij mochten spelen met de CM, zoals werd gedaan in de vijfde studie, werden verbanden gevonden met planning, probleemoplossend vermogen en werkgeheugen. Resultaten van het spel van kinderen met de CM was ook verbonden aan flexibiliteit en het nemen van initiatief zoals gemeten door de antwoorden van ouders op de vragenlijst. De bevindingen van de vierde en vijfde studie tonen aan dat er verbanden bestaan tussen de CM en EF, waarbij die laatste gemeten werd door traditionele taken en een vragenlijst.

Deze resultaten van de derde tot en met de vijfde studie laten zien dat het mogelijk om met een technologisch, 'slim' speelgoed informatie te verzamelen over de cognitieve vaardigheden of EF van kleuters. Er kan voorzichtig geconcludeerd worden dat het mogelijk is om spelenderwijs de EF van kleuters te testen, met behulp van technologische toepassingen. In de toekomst kan dit idee verder uitgebreid worden waarbij de betrouwbaarheid van dergelijke metingen verder onderzocht dient te worden. Het ontwikkelen van een reeks van slim speelgoed kan daarbij interessant zijn om vervolgens in

te zetten als meetinstrumenten. Wellicht kan een diversiteit aan speelgoed een gedetailleerd beeld van de EF van jonge kinderen opleveren.

Referenties

- Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development, 78*(2), 647-663. doi:10.1111/j.1467-8624.2007.01019.x
- Borella, E., Carretti, B., & Pelegrina, S. (2010). The specific role of inhibition in reading comprehension in good and poor comprehenders. *Journal of Learning disabilities, 43*(6), 541-552.
- Brown, T. T., & Jernigan, T. L. (2012). Brain development during the preschool years. *Neuropsychology Review, 22*(4), 313-333. doi:10.1007/s11065-012-9214-1
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*, 135-168.
- 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>
- Viterbori, P., Usai, M. C., Traverso, L., & De Franchis, V. (2015). How preschool executive functioning predicts several aspects of math achievement in Grades 1 and 3: A longitudinal study. *Journal of Experimental Child Psychology, 140*, 38-55.

About the author

Corrie Christina Jeannine Urlings was born on August 23, 1988 in Genk, Belgium. She completed her secondary education at Heilig Hartcollege Lanaken. She then started her studies in psychology at Maastricht University. Her bachelor studies were followed by a Master of Science in clinical and health psychology at the University of Leuven, which she completed with distinction.

After graduating, she worked as a clinical psychologist in second-line mental health care for children and was a team leader of a large daycare facility. In 2015 she started her academic career at Maastricht University as a junior researcher at the School of Business and Economics, department of Macro, International and Labour Economics, where she studied early childhood cognitive interventions. This position was followed by an interdisciplinary PhD project. The result of the research projects during the PhD are compiled in this thesis focusing on assessment of young children's executive functions via a compelling combination of toys and technology. During this time, she presented her work in multiple international conferences, such as EARLI (Finland 2017, SIG05 Germany 2018), EECERA (Ireland 2016), ENPAIR (Spain 2019), ORD (Netherlands 2019) and was challenge leader of in a hackathon on 'Sensors and Data' (Germany 2018). Her work has been published in international, peer-reviewed journals. She obtained a University Teaching Qualification at Maastricht University as well for her teaching experience, course development and assessment of students.

She continued her passion for assessment at Maastricht University as an educational developer in the department of Educational Research and Development in 2019. In 2020, she started her current position as an assistant professor at the Open University in Heerlen, where she teaches and conducts research at the department of Technology Enhanced Learning and Innovation.

