Combining computerized adaptive practice and monitoring: the possibilities of self-organizing adaptive learning tools

Marthe Straatemeier
Oefenweb.nl / University of Amsterdam
m.straatemeier@uva.nl

Han L. J. van der Maas
University of Amsterdam / Oefenweb.nl
h.l.j.vandermaas@uva.nl

Brenda Jansen
University of Amsterdam
b.r.j.jansen@uva.nl

Abstract

In the past few years there has been a rapid and large-scale introduction of digital technology in schools, which creates many new possibilities for educational assessment. We present a new methodology for combining web-based computerized adaptive practice and monitoring. This methodology is an extension of standard techniques of computer adaptive testing and is based on the Elo rating system (Elo, 1978). This self-organizing system enables simultaneously tracking the development of abilities and the calibration of item difficulties, as estimates of person abilities and item difficulties are updated with every answered item. In addition, both responses times and accuracy are used in the computation of ability and difficulty estimates.

This methodology was first implemented in a web-based progress-monitoring system for math (Math Garden), which was originally developed at the University of Amsterdam to meet both educational and scientific aims. Within Math Garden children practice their math skills by playing games with items matched to their ability level. At the same time, their progress is automatically being monitored and presented to their teachers. In this paper we will give an overview of the underlying ideas and discuss the research that has been carried out with the Math Garden dataset.

Key words: computer adaptive testing, high frequent monitoring, web based, educational assessment

Introduction

The measurement of children's cognitive abilities and development is important in both educational and research settings. To measure and understand the complex processes that play a role in children's cognitive development a time-serial approach is needed. For this aim we developed Math Garden (www.mathsgarden.com). Math Garden is a web application in which children can play math games. The garden metaphor is used to stimulate children to maintain their mathematical abilities as they can nurture their plants by improving their mathematical ability and prevent their plants from withering by playing on a regular basis. Math Garden uses a new method for computer adaptive testing by which the difficulty of the exercises or items is automatically adjusted to the skills of the child.

Our vision on the use of Math Garden in an educational setting can best be described by the concept of Game, Train, Track and Teach. Within Math Garden children play adaptive **games** in a stimulating online environment and thereby **train** their mathematical skills on their own ability level. The data of these training sessions are **tracked** as the answers and response times of every solved problem are being registered in an online database. This database enables comparisons of pupils and school classes to their reference groups, informing teachers about strengths and weaknesses of their pupils. In addition, information about children's specific errors and strategies can be given. Teachers can use this information to optimize **teaching** at both the individual and school class level. With Math Garden we aim to take over the less appealing parts of education, that is, many hours of practice on the student level and correction of school work on the teacher level, making them more pleasant and integrating them with the aim of monitoring children over time.

Before introducing the instrument in more detail, we review a number of key ideas that have influenced the development of the Math Garden system.

Mathematical development is a complex dynamical system

We view cognitive learning and development as a complex dynamical system. Complex dynamical systems are considered to be a network of many elements that interact iteratively with each other (van der Maas, Dolan, Grasman, Wicherts, Huizenga, & Raijmakers, 2006; van der Maas & Molenaar, 1992). They develop and adapt dynamically and are often characterized by non-linearity and self-organization. We argue that an important way to study children's complex system of interacting abilities is the microgenetic approach (Siegler & Crowley, 1991), that is, high frequent measurements should be used. Researchers using the microgenetic method are specifically interested in the process of change, not just in the product of change (Granott & Parziale, 2002). The microgenetic method is characterized by high frequent measurements during a period of development. The density of these measurements should be high relative to the rate of change and the study should span the whole developmental process, that is, until a relatively stable state is reached. In addition, the collected data should be analyzed with intensive trial-by-trial analyses in order to detect the dynamics of the developmental processes (Flynn, Pine & Lewis, 2006; Siegler & Crowley, 1991). With this method key features of developmental processes can be detected, such as transitions, sensitive periods, but also relapses and stagnations (van der Maas, Jansen, & Raijmakers, 2004; van der Maas & Molenaar, 1992; van der Maas & Raijmakers, 2009).

These findings could also have a large effect on education. If, for example, we are able to detect certain sensitive periods in which children are especially susceptible to specific instructions, more effective instructions can be given.

Mathematical abilities can be considered as a form of cognitive expertise

Our second key idea is that children’s arithmetical development and learning should be seen as a form of expertise development. With expertise we mean the outstanding performance within a certain domain, for example sports, music or science. Ericsson, one of the leading experts in the domain of expertise development, claims that motivation and practice are the key principles in expertise development (Ericsson, 2006; Ericsson & Ward, 2007). According to Ericsson the development of expert performance is gradual and expertise is only acquired after many years of special practice activities, which he calls deliberate practice. Deliberate practice is characterized by goal-directed training with repeated exercises just beyond the current ability level and with immediate feedback. Because of the intensity of these goal-directed training sessions regular sessions are preferred over a small number of long training sessions (Ericsson, 2006). In addition, Krampe and Charness (2006) argue that deliberate practice is also necessary for maintaining expert performance.

To optimize mathematics education one should strive to implement the principles of deliberate practice into everyday education. Dutch policymakers also argue that more time should be dedicated to practice in everyday education (Expert group “Doorlopende leerlijnen, 2008). The Dutch ministry of education acknowledges the importance of maintaining and extending one’s knowledge with age by implementing basic math (i.e., arithmetic) exams at several time points in higher education. We believe that there is much to win if practice in education is implemented according to the principles of deliberate practice. However, one of the main requirements of deliberate practice is intensive one-on-one guidance by a teacher or coach. A classroom of 30 children, instructed by one teacher stands in clear contrast to the individual coach of promising athletes or musicians. Bloom (1984) also claimed that one-to-one tutoring is very beneficial and refers to this effect as the two sigma problem. He found that students who received one-to-one tutoring performed on average two standard deviations better than students who received conventional instruction within a classroom setting. Providing every student with the most optimal learning settings according to the principles of deliberate practice is extra complicated by the individual differences that exist within a classroom, which brings us to our third key idea.

Individual differences in Math are huge

Our third idea is that individual differences in mathematical ability are huge. In the Math Garden dataset we found large differences in mathematical ability in all grades. Figure 1 shows the proportion of children per grade that score above or below the mean of one or two grades higher and lower, averaged across addition and subtraction. In all grades there is a substantial number of children who score above the mean of two grades higher: between 6.5% to 14.9%. On the other hand there are also many children, between 7.0% to 26.5%, who score below the mean of children in two grades lower. Results from the periodical educational assessments performed in the Netherlands by CITO also demonstrated considerable individual differences in math ability within Grade 3 (Hop, Janssen, Hemker, van Weerden & Til, 2012) and Grade 6 (Scheltens, Hemker, & Vermeulen, 2013). These results illustrate the enormous challenges that teachers face when teaching 30 children with varying abilities.

These large individual differences make it especially difficult for teachers to provide optimal practice sessions for each child. At the same time the Dutch ministry of Education demands that all children, also children who are weak or excelling in math, receive education at their own level. One of the prerequisites of good adaptive education is knowledge about the ability levels of pupils. One can only provide suitable instruction if it is known at what level a child functions and what his or her strengths, flaws, and common errors are. Only then it is possible to provide instruction according to the principles of deliberate practice. Many schools in the Netherlands use progress-monitoring systems (Blok, Otter, & Roeleveld, 2002) but most method-independent progress-monitoring systems, such as the child monitoring system of Cito (Janssen, Verhelst, Engelen & Scheltens, 2010), measure children's abilities only once or twice every year. To be able to detect problems early and start remediation in time, measurements should take place more frequently. However, the more educational time is used for testing, the less time remains for practice and instruction.

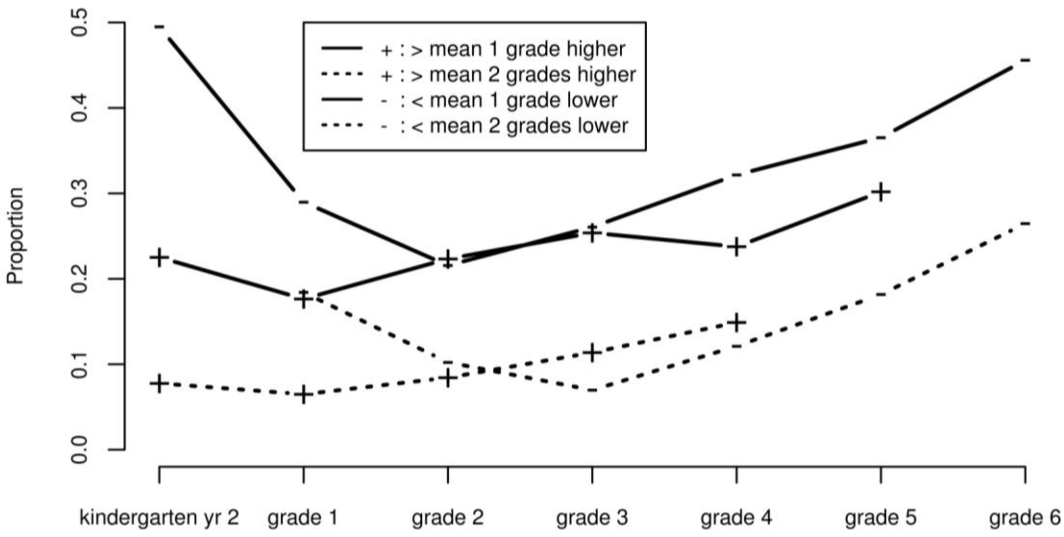

Figure 1: Proportion of children scoring above the mean of children in one or two grades higher and lower, averaged across the domains addition and subtraction, in the months April till June 2013.

ICT is sufficiently developed to enable optimal practice activities and measurement in classroom settings

To summarize, in order to develop an instrument that meets both educational and research goals, we need an instrument that measures children's math abilities on a high frequent basis, can meet the individual differences in ability and implements the principles of deliberate practice. The solution can be found in ICT. In the past few years, we have witnessed the rapid and large-scale introduction of computer technology in households and schools, due to the availability of small computers (mini laptops, handhelds) and fast WIFI access to the internet. According to Statistics Netherlands (2011) 92% of the Dutch households had a computer in 2010 and 91% of the households had internet access. Moreover, the increasing use of smartphones and tablets has lead to strong increase in the volume of internet traffic. Developing educational instruments that exploit these new possibilities is a major challenge, which is being taken up by many research groups, companies, and schools.

Math Garden

With these ideas in mind we developed Math Garden. Math Garden is a web based training-tracking system in which children can train their mathematical skills while at the same time their development is being measured. After logging in, children enter their personal garden in which every plant represents a math game. Several game principles are implemented to motivate children to practice their math abilities on a regular (weekly) basis.

The state of the garden gives children an indication of their mathematical skills as the size of the plants indicates their ability level. Plants grow bigger when their ability increases. If children do not maintain their garden by playing regularly, the plants will wither as a cue that these math skills need to be practiced. In September 2013 Math Garden contains sixteen games measuring math-related skills such as counting, number series, telling time and fractions as well. Each game administers items from a large item bank (> 400 items per game). These item banks consist of items, varying in difficulty, representing the curriculum of primary education and beyond.

Within Math Garden each child receives items adjusted to his or her ability level. This was accomplished by using a new adaptive testing procedure, using state of the art psychometric modelling. One of the key principles of this adaptive system is self-organization. Firstly, this applies to the computer adaptive testing (CAT) system that is used to administer items in the math games. One of the requirements of an adaptive system is that the difficulties of the items are known on beforehand. Adjusting item difficulty to the ability level of a child, demands knowledge on the difficulty of items. This requires pre-testing of items, which is expensive and time consuming. CAT is mainly used by large companies and applied in large-scale educational settings. The new CAT system in Math Garden makes the process of pretesting unnecessary. It is based on the Elo rating system that has been developed to compare chess players (Elo, 1978). Comparable rating systems are used in various sports and games to measure the ability of players and to match opponents of equal strength.

In Math Garden children and items are considered opponents and thus a child solving an item is seen as a match (Klinkenberg, Straatemeier, & van der Maas, 2011). Within this system both items and children have a rating, indicating their difficulty or ability level, respectively. After each solved item the rating of the child and the rating of the item are adjusted. For example, if a child answers an item incorrectly, the item wins. The item gains ratings points, that is, becomes more difficult, and the child loses rating points. The amount of rating points won or lost depends on the difference in rating between the child and the item. This new CAT system and the online setup of Math Garden enable the estimation of both item difficulties (item ratings) and children's abilities (person ratings) on the fly. Estimations of item difficulties are, therefore, based on the answers of all children playing online in Math Garden. This self-organizing nature of the CAT system enables the fast and easy development of new adaptive games. After an item set is constructed with items of various difficulties, the CAT system will do the rest. A schematic overview of the Math Garden system is presented in Figure 2.

Both the accuracy and speed of children's answers are combined to estimate their mathematical ability with a new scoring rule: the High Speed High Stakes (HSHS) scoring rule (Maris & van der Maas, 2012). For each item children have limited time to give an answer. After an answer is given the score on an item equals the remaining time and this score is either positive, in case of a correct answer, or negative, in case of an incorrect answer. This scoring rule is implemented in the games in a playful manner: The deadline for an item is visualized by coins on the screen and every second a coin disappears. Children either win or lose the remaining coins when providing a correct or incorrect answer. Children can use earned coins to buy trophies for their virtual trophy cabinet. Using the extra information of response times enables the administration of easier items than those that are used in standard computer adaptive tests. In standard computer adaptive tests items are selected for which the child has a probability of 50% of answering correctly. These items are the most informative items for measuring a person's ability but lead to a, possibly demotivating, high level of negative feedback. In Math Garden, we present items with a probability of answering correctly of 75% (or even higher) and use the extra information of the response times to estimate children's abilities.

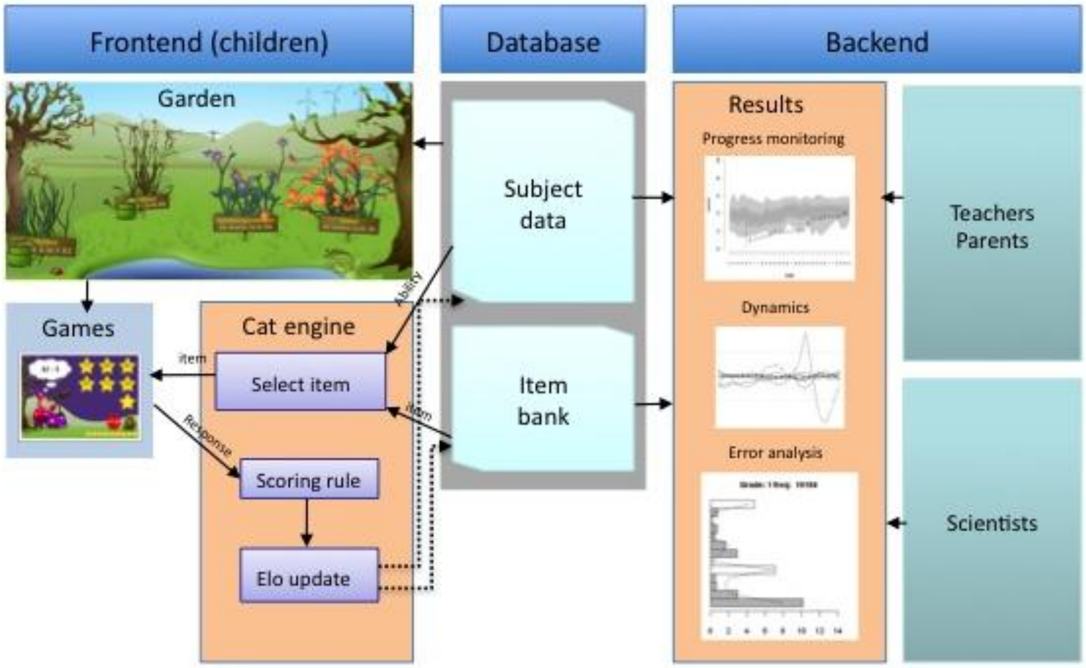

Figure 2: A systematic overview of the web application Mathsgarden.com (Rekentuin.nl). Children (frontend) can maintain their garden by playing math games. The size of the plants depends on the ability level of the child. Item choice is adapted to the ability of the child by the cat engine. Estimates of children's ability and item difficulty are updated after every answered item according to an extended Elo algorithm with the HSHS scoring rule. Teachers, parents and scientist receive automatically generated reports on the data.

Klinkenberg et al. (2011) showed that the Elo system combined with the HSHS scoring rule resulted in less bias and higher measurement precision for easy items compared to standard CAT. Thus, our new adaptive technology enables the administration of easy items while still effectively measuring ability. Moreover, they also showed that both the item difficulty estimates of the items and the ability estimates of the children are reliable and valid. With the current participant sample the item ratings will converge to stable item ratings in a short period of time. This makes Math Garden a platform for the fast development of new adaptive games. When a new item bank is available it can be implemented in a Math Garden game and the item bank is calibrated within a few days. Other indications of high reliability of the item ratings are the high correlations of the item ratings and discrimination parameters between sets of parallel items (Klinkenberg et al., 2011). Evidence for the reliability and validity of the person ratings was found in the fairly high correlations (range: .67 - .88) between the domains addition, subtraction, multiplication, and division. Other results supporting the validity of the person ratings are the high correlations between person ratings on these four domains and the general Math ability scale of the pupil monitoring system of Cito (Janssen & Engelen, 2002) and the positive relation between grade and person ratings.

Another aspect of self-organization of the Math Garden system concerns the reports about children's performance, discussed above. One of the important goals of testing in education is the comparison of children's performance to their norm group, which enables the detection of children that are ahead or behind and need extra attention. The construction of norm groups is, however, a very costly and time-consuming process. It requires a large representative sample of children in each age group. Moreover, norm groups may become out-dated as the educational curriculum changes and they are often bound to certain time points of the school year. The online set-up of Math Garden enables the comparison to other users at every time point, as all data is stored in a central database. We speak of reference groups instead of norm groups because these groups have not been formed according to standard norm requirements. For example, there is no control over the type of schools that use Math Garden. Also, the conditions under which children use Math Garden are less controlled than the conditions under which norm-referenced tests are being administered. Children can use Math Garden anytime anywhere as long as there is an internet connection. Some requirements are used to ensure the reliability of the reference groups: only children that have played sufficient items in a recent time frame are included.

The advantage of the reference groups in Math Garden lies, however, in the power or amount of data. The more users play the games in Math Garden the more reliable the reference groups become. For example, the norm groups in the monitoring system of Cito for Math consist of 778 to 1516 users in each age group (Janssen et al., 2010). In September 2013 Math Garden has between 11.000 to 17.000 active users in grade 1 to 6. We consider Math Garden an example of low-stakes-testing. The reference group comparison can give a fairly good indication of children's performance at every desired time point. Moreover the online set-up allows for self-organization of the reference groups. Changes in performance across the school year, either developmental or due to changes in the curriculum, will be reflected in the performance of the reference group because the performance of all users is tracked over time. Summarized, Math Garden does not have the disadvantage of the need of repeatedly conducting new research for determining the norms that off-line tests and online tests that use static norm groups do have. The larger the group of Math Garden users, the more specific the reference groups can be. We could then, for example, compare 8-year old girls who attend a 'Montessori'-school in the region Amsterdam with each other. International use of Math Garden would allow assessment of international differences in math performance, possibly as an alternative for large expensive research projects such as TIMMS and PISA.

This brings us to the final characteristic of Math Garden we would like to discuss: Math Garden combines practice and testing in one program. Within the educational field there is a clear distinction between practice and testing. Much time can be saved if these two goals are combined. This is what Math Garden does. Math Garden uses children's daily practice to track their performance, using psychometric modelling. Why administer expensive time-consuming tests when children's everyday practice sessions provide a rich dataset concerning their performance? Again the advantage lies in the power of data. The introduction of tablets and small computers in education enables children to practice in Math Garden on a weekly or daily basis, by which their performance is automatically tracked. The more frequently children use Math Garden, the better the insight in children's math development. In addition, a bad day or help from a sibling or parent can easily be detected when comparing a child's performance to his or her performance on other occasions. Thus, Math Garden enables high-frequent monitoring of children's performance while at the same time children practice according to the principles of deliberate practice. Thereby, no valuable time is lost on the administration of tests.

The success of the Math Garden project has led to the start-up of Oefenweb.nl. Oefenweb.nl is a spin-off company of the University of Amsterdam aimed to improve Math Garden for educational use and to develop other online learning tools based on the same principles. Nowadays, users of Math Garden are still asked for permission for use of their data for scientific research, thereby maintaining the link between education and research. In November 2012 the Language Sea (NL: Taalzee) has been launched. Language Sea is similar to Math Garden: it is an online learning environment for practicing and measuring language skills such as grammar, reading, and vocabulary. Two other applications concern touch typing and learning statistics.

Discussed so far is how Math Garden meets the educational aim of providing optimal learning and measurement settings for children of all abilities. The data of Math Garden is also used for scientific research at the University of Amsterdam. In the past 5 years, we collected a large and unique longitudinal data set (over 230 million responses by children of more than 800 schools) for educational and developmental researchers. We will, therefore, end this paper with an overview of the research possibilities using the Math Garden and the resulting dataset.

One line of research concerns the analysis of item difficulties of the item banks of Math Garden. Several studies focus on explaining the item difficulties by a limited set of item characteristics. Insights into the item characteristics affecting item difficulty and the interactions and dependencies between these characteristics can give us insights into the challenges children face when mastering mathematical problems. Using this information in the development of math methods may improve education. For example, Straatemeier, Jansen, and van der Maas (submitted) showed that a limited set of item characteristics could explain a large amount of variance of addition and subtraction problems (86% - 91%). They concluded that the most robust increasing effect on item difficulty was whether carrying (addition) or borrowing (subtraction) was required to solve the problem. Other studies have focused on the item difficulties of multiplication problems (van der Ven, Straatemeier, Jansen, and van der Maas, submitted); items of a task for logical reasoning, a deductive version of the well-known Mastermind game, (Gierasimczuk, van der Maas, & Raijmakers, in press); item difficulties of enumeration problems (Jansen, Hofman, Straatemeier, van Bers, Raijmakers, & van der Maas, submitted); and items of a visuospatial working memory task (van der Ven, van der Maas, Straatemeier, & Jansen, submitted).

Another line of research concerns the effect of adaptive learning tools on children's motivation. The results of the study of Klinkenberg et al. (2011) support our assumption that a CAT procedure with easy items has a positive effect on children's motivation. Children played a lot outside school hours (33.2 % of the answers) and children with low ability did not play appreciably less, indicating that motivation was similar for children of all abilities. In another study we further investigated the relation between Math Garden, motivation, math anxiety, perceived math competence, and math performance. The success rates of children were set at .6, .75, and .9. We found that the higher the success rate in Math Garden, the more children played the Math Garden games and the larger the improvement in math performance. We conclude, therefore, that the experience of success stimulates practice and that practicing math frequently at one's own ability level improves math performance (Jansen et al., 2013). The latter conclusion is also supported by the study of Jansen, De Lange and Van der Molen (2013) who found that adolescents (12-15 years) from special education who frequently played in Math Garden improved more in math ability than the control group who did not play in Math Garden. The sample size in the study was, however, small.

In the normal setup of Math Garden children are free to choose between the three difficulty levels (success rate of .6, .75, or .9), allowing children to control the frequency of negative feedback. Hofman, Jansen, Visser, and Van der Maas (submitted) studied children's choices and found that children in higher grades and with higher ability levels choose difficult items more often than children in lower grades and with lower ability levels. In addition, boys tend to prefer more difficult items than girls. These results indicate that there are individual differences in preference for difficulty levels and that self-adaptive training and testing may be a valuable feature in individual training programs such as Math Garden.

Currently research projects at the University of Amsterdam focus on a number of questions regarding the dynamics of development, namely analyzing developmental change with the time-series data of Math Garden and regarding the mutual relations between cognitive and scholastic abilities. For example, Van der Ven, Van der Maas, Straatemeier and Jansen (submitted) found that visuospatial working memory and mathematics are significantly related, but that this is especially true for the domains addition and subtraction in the lower grades. For the domains multiplication and division this relationship was weaker and no age trend was found. This study illustrates that Math Garden allows studying interrelationships of a large range of abilities with children of a large age range.

References

Blok, H., Otter, M. & Roeleveld J. (2002*).* Leerlingvolgsystemen: Praktische lessen uit onderzoek, *[Student monitoring systems: Practical lessons from research]*. Amsterdam, The Netherlands: SCO-Kohnstamm.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13,* 4-16.

Commissie Dijsselbloem, 2008. Parlementair onderzoek onderwijsvernieuwingen, *[Parliamentary inquiry educational reform]*, Sdu publishers, The Hague (2008).

Elo, A. (1978). *The Rating of Chessplayers, Past and Present*. London: Batsford.

Ericsson, K. A. (2006). The influence of experience and deliberate practice on the development of superior expert performance. In: Ericsson K., Charness N., Feltovich P., Hoffman R. (Eds.), *Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press, 723-742.

Ericsson, K.A., & Ward P. (2007). Capturing the naturally occurring superior performance of experts in the laboratory: Toward a science of expert and exceptional performance. *Current Directions in Psychological Science, 16,* 346-350.

Expertgroep doorlopende leerlijnen, 2008. Over de drempels met taal en rekenen, *[crossing the barriers of language and mathematics]*, Nationaal expertisecentrum leerplanontwikkeling.

Flynn, E., Pine, K., & Lewis, C. (2006). The microgenetic method: Time for change? *The Psychologist, 19*, 152-155.

Gierasimczuk, N., van der Maas, H. L. J., & Raijmakers, M. E. J. (in press). An analytic tableaux model for deductive mastermind empirically tested with a massively used online learning system. *Journal of Logic, Language and Information.*

Granott, N. & Parziale, J. (Eds.) (2002). *Microdevelopment: Transition processes in development and learning.* Cambridge, UK: Cambridge University Press.

Hofman, A. D., Jansen, B. R. J., Visser, I., and Van der Maas, H. L. J. (submitted). Individual differences in Self-Adaptive Testing in Math and Language.

Hop, M., Janssen, J., Hemker, B., van Weerden, J., & van Til, A. (2012). Balans van het reken-wiskundeonderwijs halverwege de basisschool 5, *[Fifth assessment of mathematics education halfway primary school]*. Arnhem, The Netherlands: CITO.

Jansen, B. R. J., Hofman, A., Straatemeier, M., van Bers, B., Raijmakers, M. E. J., & van der Maas, H. L. J. (submitted). The role of pattern recognition in children's exact enumeration of small numbers.

Jansen, B. R., Louwerse, J., Straatemeier, M., Van der Ven, S. H., Klinkenberg, S., & Van der Maas, H. L. (2013). The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. *Learning and Individual Differences, 24*, 190-197.

Jansen, B. R., De Lange, E., & Van der Molen, M. J. (2013). Math practice and its influence on math skills and executive functions in adolescents with mild to borderline intellectual disability. *Research in developmental disabilities, 34*, 1815-1824.

Janssen, J., & Engelen, R. (2002). Verantwoording van de toetsen Rekenen-Wiskunde, *[Justification of the Cito student monitoring system for arithmetic performance]*. Arnhem, The Netherlands: CITO.

Janssen, J., Verhelst, N., Engelen, R., & Scheltens, F. (2010). Wetenschappelijk verantwoording van de toetsen LOVS Rekenen-Wiskunde voor groep 3 tot en met 8, *[Scientific justification of the Cito student monitoring system for arithmetic performance for grade 1 to 6]*. Arnhem, The Netherlands: CITO.

Klinkenberg, S., Straatemeier, M., & Van der Maas, H. L. J. (2011). Computer adaptive practice of maths ability using a new item response model for on the fly ability and difficulty estimation. *Computers & Education, 57*, 1813-1824.

Krampe, R., Charness, N. (2006). Aging and expertise. In: Ericsson K., Charness N., Feltovich P., Hoffman R. (Eds.), *Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press, 723-742.

Maris, G. & Van der Maas, H. L. J (2012). Speed-accuracy response models: Scoring rules based on response time and accuracy. *Psychometrika, 77,* 615-633.

Scheltens, F., Hemker, B., & Vermeulen, J. (2013). Balans van het reken-wiskundeonderwijs aan het einde van de basisschool 5, *[Fifth assessment of mathematics education at the end of primary school]*. Arnhem, The Netherlands: CITO.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46*, 606-620.
Statistics Netherlands (2011). *ICT, Knowlegde and the Economy.* The Hague, The Netherlands, Statistics Netherlands.

Straatemeier, M., Jansen, B. R. J., & Van der Maas, H. L. J. (submitted). Difficulties of simple and complex addition and subtraction problems.

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P. , Wicherts, J. M. , Huizenga, H. M. & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842-861.

van der Maas, H. L. J., Jansen, B. R. J. & Raijmakers, M. E. J. (2004). Developmental patterns in proportional reasoning. In A. Demetriou & A. Raftopoulos (Eds.), *Cognitive Developmental Change Theories, Models and Measurement*. Cambridge University Press.

Van der Maas, H.L.J., & Molenaar, P.C.M. (1992). Stagewise cognitive development: an application of catastrophe theory. *Psychological Review*, *99*, 395-417.

Van der Maas, H. L. J., & Raijmakers, M. E. J. (2009). Transitions in cognitive development: prospects and limitations of a neural dynamic approach. In J.P. Spencer, M. Thomas, & J. McClelland (Eds). *Toward a New Grand Theory of Development: Connectionism and Dynamic Systems Theory Re-Considered*. (p.299-312). Oxford University Press.

Van der Ven, S., Straatemeier, M., Jansen, B. R. J. & Van der Maas (submitted). The difficulty of single digit and multidigit multiplication problems and performance at different ages in primary school.

Van der Ven, S., van der Maas, H. L. J., Straatemeier, M., & Jansen, B. R. J. (submitted). Visuospatial working memory and mathematical ability at different ages throughout primary school.