# The development of theory-of-mind and the theory-of-mind storybooks

Blijd-Hoogewys, Els Maria Arsene; Blijd-Hoogeweys, E.M.A.

# CHAPTER 6

## Summary and general discussion

**Theory-of-Mind research**

Theory-of-Mind (abbreviated as ToM) refers to the ability to attribute mental states – such as beliefs, desires and intentions – to oneself and others and to use these attributions in understanding, predicting and explaining behavior. It is a social cognitive ability which typically developing children develop roughly between their third and sixth birthday.

To assess ToM ability in children, researchers often use single task measurements, involving single aspects of ToM, most commonly the comprehension of false belief, like the Smarties test (Perner et al., 1987) and the Sally and Anne test (Baron-Cohen et al., 1985). These assessments may be quick and efficient, but provide little information about the bigger construct of ToM and its developmental pathway over time. ToM is a complex ability that comprises far more than the understanding of false beliefs, and includes for instance also the understanding of desires and emotions. Therefore, more comprehensive ToM tests are advisable, in order to get more insight in the development of ToM. In fact, research has shown that comprehensive ToM tests using compound scores are more stable, because they average over multiple factors and lead to a more accurate measurement of the underlying skill (Hughes et al., 2000). In using such scores, a more adequate diagnostic procedure is attained, which can help in studying the potential nature and causes of ToM differences and ToM problems in children (Hughes & Dunn, 1998).

A well-known group of children with severe ToM problems is children with an autistic disorder (for an overview, see Baron-Cohen, 1989b, 2000). Less is known about the ToM functioning of children with lesser variants of autism, like children with PDD-NOS (Pervasive Developmental Disorder Not Otherwise Specified). Little research is devoted to this clinical group, in spite of the fact that they form a much larger group within population and within clinical practice than children with an autistic disorder. Therefore, the goal of this dissertation was to get insight in the development of ToM in children with PDD-NOS. Since so little is known, we decided to focus on the development of early ToM abilities. Early ToM abilities are those abilities typically developing children acquire between their third and fifth year of life.

It was stipulated that a design using multiple repeated measurements is the best procedure in order to follow the ToM development over time. In order to do so, an instrument was needed that could assess early ToM abilities, that could pinpoint the specific ToM problems of children with PDD-NOS in comparison with typically developing children, and that was

applicable in a longitudinal design. At the onset of this dissertation, such a test was not available. *Hence, the first objective of this dissertation was to construct a comprehensive ToM instrument, incorporating different relevant ToM tasks, applicable for a wide age range in both typically developing children and children with PDD-NOS*. The test was called the ToM Storybooks. Alternative versions were developed in order to enable repeated measurements. These alternative versions are based upon the same task structure but consist of different task content so explicit recognition of identical items would not occur.

In order to enable comparisons between children with PDD-NOS and typically developing children, *the second objective of this dissertation was to determine the psychometric qualities of the test and to calculate norm scores for ToM as a whole and for its different components*. Following the validation and norming of the ToM Storybooks, *the third objective was to describe the development of ToM in typically developing children*. Finally, *the fourth objective was to map the development of ToM in children with PDD-NOS*.

The group of typically developing children consisted of 324 subjects between 3 and 12 years of age, who were tested only once. The group of children with PDD-NOS comprised 30 children between the ages of 3 and 8 years, followed over a period of 20 months, and were tested every 4 months. All statistics were based on random permutation techniques, and more generally, Monte-Carlo analyses, which are assumption-free techniques. Wellman and colleagues (2001) have argued for the use of more assumption-free techniques, such as bootstrap methods, in ToM research. These techniques entail a simulation of the test statistic at issue as based on the null hypothesis, which can be compared to empirical ToM data (Good, 2001; Manly, 1997; Todman & Dugard, 2001). They are highly flexible and robust methods.

## Construction of the ToM storybooks

Chapter two focuses on the first and part of the second objective of this dissertation, which is the construction and validation of the ToM Storybooks. We wanted to develop a comprehensive ToM test that assesses a variety of early ToM components and associated aspects, applicable for a wide age range in both typically developing children and children with PDD-NOS. In deciding which aspects to incorporate in the test, we relied primarily upon the work of Wellman (1990), not only focusing on core ToM

components, like desires and beliefs, but also on associated aspects, like the recognition of emotions, perception knowledge and the difference between physical and mental entities. In the end, five subtypes of abilities were aimed at, namely emotion recognition, understanding of desires and beliefs, making the distinction between physical and mental entities, and seeing leads to knowing. These aspects should be completely developed in typically developing five-year olds.

Different tasks were selected and formulated in Dutch (for example tasks, see Appendix A). This resulted in a total of 34 tasks consisting of multiple questions. Because there were so many tasks, different procedures were followed to keep the testing fun and reliable. First of all, the tasks were assimilated in six story lines (six storybooks), so children could more easily follow the tasks and would not get tired of testing. Next to that, each task was illustrated with a full color picture. Also transitions between tasks received accompanying drawings and text, to keep the story going and to avoid too much switching between tasks. Furthermore, the drawings were enlivened by the use of toy doors that can be opened, magnetic emotion faces that can be placed on the characters, and patches of soft fur that can be caressed, if wanted. And finally, children could choose the order in which four of the books were read to them (this applied to books 2 to 5, which have similar underlying structures), aimed at increasing the child's commitment and motivation.

There are 95 questions in total, including 18 justification questions. Each answer receives zero or one point (correct versus incorrect), unless it involves a justification. In the latter case, zero to two points is awarded, dependent on the correctness and quality of the justification in question. In order to score the justifications in a standard fashion, a category system was developed. This was based on the category system used by Rieffe (1998), on different categories from Wellman (1990), and on an exploration of the empirical data. The result was a system which could reliably categorize answers (see Appendix B & C). The inter-rater reliability for this system was high (Cohen's Kappa =.81-.97; based upon five independent raters).

After omission of psychometrically inadequate items, 92 items remained. The answers resulted in a compound ToM sumscore with a maximum total score of 110 points (ToM-total score). A one-parameter logistic model (OPLM) based upon this sumscore showed a good fit. Thus, the ToM sumscore can be considered a good estimation of the ToM ability of a child.

Over a period of one and a half years, the ToM Working Group (Blijd-Hoogewys, Van Geert, Loth & Serra) developed four different versions of the ToM Storybooks: version Sam, version Lotje, version Pieter and version Hanna. Earlier, a preliminary version was developed by Van Geert (drawings), Serra and Loth (story lines) (reported on in Serra et al., 2002). Based on this preliminary version, the current versions were developed[1]. Concerning the versions Sam and Lotje, the story lines were developed by Serra, Loth and Blijd-Hoogewys. Concerning the versions Pieter and Hannah, the story lines were developed by Serra and Blijd-Hoogewys. The accompanying drawings were made by Blijd-Hoogewys and van Geert. The author of this dissertation subjected each version of the ToM Storybooks to pilot studies, undertaken in both typically developing children and children with autism spectrum disorders. Based upon the results of these pilot studies, alterations were made to the text or drawings and subsequent pilot studies were initiated.

Results showed that the ToM Storybooks have good psychometric qualities (for a quick overview, see Table 9 in Chapter 2, pp. 44). A component analysis resulted in a structure that largely corresponds with the underlying theoretical constructs from the test, affirming content validity. The internal consistency of the ToM Storybooks was good ($\alpha$=.90). The test-retest reliability in both typically developing children and children with PDD-NOS was good (r=.86 and r=.98 respectively). The short term (after 2 to 3 weeks) test-retest correlations in typically developing children suggested a learning effect, consistent with other test-retest researches (such as with intelligence tests). This was not seen in children with PDD-NOS. The inter-rater reliability was high (Cohen's Kappa =.97-.99). No nuisance

[1] The following major adjustments were made. First, we have adjusted the drawings by applying a more coherent drawing style (for instance, the characters always wear the same cloths, and unnecessary details were removed). Second, the different tasks follow one another more naturally; and transitions between tasks are more often accompanied by additional drawings and text. Third, in order to keep the children motivated and interested, the drawings were enlivened by the use of non-graphical elements that are distributed sparsely across the text. They include patches of soft fur, toy doors that can be opened, and magnetic emotion faces that can be placed on the characters. The manipulation of these elements is not a necessary condition for answering the test. We like to point out that none of the children from the clinical populations to which the test has been administered so far has shown any sign of disturbance or aversion for the non-graphical elements. Fourth, based on item- and factor-analyses, items have been removed or changed (for instance, in the emotion recognition tasks the emotion 'curious' was removed and the emotions 'anger' and 'fear' were added).

effects were found. The discriminant validity was good. The ToM Storybooks are able to identify children with ToM problems, in that the scores of children with PDD-NOS were significantly lower than those of matched control children. Convergent validity was good: the test displayed significant correlations with other tests as suspected. There were high correlations with the ToM test (r=.79), negative correlations with the CSBQ (a.k.a. the 'VISK'; the lower the ToM score, the more autism-related behavioral problems reported by parents), and positive correlations with the VABS questionnaire (a higher ToM score implies higher interactive sociability). Divergent validity was as expected: ToM scores correlated with language comprehension and intelligence.

One of the methodological strengths of the current test is that it has extended the limitations common in the majority of the researches done in the field of ToM. Most research has been undertaken in young children only (mostly up to 6 years, with a major focus on 3 to 4 year olds), has used only a few tasks (mainly false belief tasks) and considered small research groups. This dissertation, aimed at constructing a new test, used a wide range of tasks (not only false belief tasks) and consisted of a substantial number of children over a wide age range.

It can be concluded that the validity and reliability of the ToM Storybooks comply with the requirements of an instrument of its sort. The psychometric values found are consistent with findings from comparable research on standard and complex false belief tasks (Hughes et al., 1999, 2000; Muris et al., 1999; Wellman et al., 2001). Ergo, the first and part of the second objective of this dissertation were achieved successfully. That is, constructing and validating a comprehensive ToM instrument, incorporating different relevant ToM tasks, applicable for a wide age range, applicable in both typically developing children and children with PDD-NOS, and deployable in making comparisons between both groups.

**Calculating norm scores for the ToM storybooks**

Chapter three is dedicated to the second part of the second objective, which is the calculation of norm scores for the ToM Storybooks. Such scores are needed in order to compare children and to evaluate inter-individual differences. We decided to calculate quotient scores, equivalent to intelligent scores, with an average of 100 and a standard deviation of 15. Such scores are well-known norm scores for practitioners. The ToM Quotient score is referred to as ToMQ.

The norm scores were calculated based on the test administered as a complete design - in which all children received all tests - instead of as a staircase design - in which a child has to succeed on a preceding module in order to be presented with the next module. The latter is based on the idea that ToM develops in a consistent manner with successive progressively developing ToM abilities, which we believe is doubtful taking into account developmental discontinuities.

Despite the fact that the ToM Storybooks contain ToM items typically developing children develop between their third and fifth year of life, we constructed norm scores for a wider age range, namely up to 12 years old. Such norms will enable us to employ the test with children who have ToM problems and whose ToM development is likely to be delayed.

Traditionally, norm scores are obtained on the basis of curve fitting of the raw data using regression analysis. Such curves are monotonically rising, not taking into account possible developmental regressions. The latter are even considered unwanted. However, sudden jumps or plateaus are recurring and well-known features in many developmental abilities. They are thought of as specific properties of developmental progress. They should also be taken into account in calculating norm scores. Therefore, we introduced a different way of calculating norm scores, based upon a non-linear smoothing method of the raw data. This is the Loess smoothing method, which stands for locally weighted least squares estimates. By taking into account the weights of neighbouring observed scores, the resulting smoothed curve more faithfully follows plateaus and temporary developmental regressions in the raw scores.

In order to validate the use of the Loess method, we compared it with a monotonically rising regression model. Because gender differences were obvious in the youngest and oldest children, separate norms for boys and girls were computed. Based on Z-scores, ToMQ scores were calculated, in addition to confidence intervals and age equivalents.

It could be concluded that the Loess smoothing procedure was superior over the monotonically rising regression model. First of all, only the Loess model could reveal differences between boys and girls (see Figure 2 in Chapter 3, pp. 62). Next, the normality assumption could not be kept for the monotonically rising regression model. The Loess method provided a considerably better approximation to the normal distribution for each age segment. Third, the average deviation from the average of 100 and the standard deviation of 15 were higher for the monotonically rising regression model in comparison with the Loess model. And fourth, confidence

intervals illustrated that the Loess model approached the average of 100 more consistently than the monotonic model (see Figure 4 in Chapter 3, pp. 65).

With this chapter, objective two of this dissertation is completely achieved. The psychometric qualities are good (see Chapter two) and norm scores are calculated. As a result, children with differing ToM levels can be compared and the distribution of ToM can be established over a wide age range. The next two chapters pursue these objectives, with Chapter four focussing on the ToM development in typically developing children, and Chapter five in children with PDD-NOS.

**ToM development in typically developing children**

Chapter four coincides with objective three of this dissertation: mapping the ToM development in typically developing children. The principal point-of-departure is that ToM development need not be gradual and continuous, but is expected to evidence developmental discontinuities, such as temporary accelerations, decelerations, delays and regressions. Such temporary discontinuities are well-known characteristics in developmental psychology and are thought of as hallmarks of underlying developmental transitions. Preceding the mastering or acquisition of a new ability, children can show developmental regressions. Children evidence a temporary relapse before their ability consolidates.

In order to explore eventual temporary discontinuities, a quantitative approach of the data was adopted. This refers to an approach that defines a quantitative dimension, for instance the level of false belief understanding on a measurement scale, and which studies the changing level of that dimension over the course of developmental time. The research was less aimed at qualitative aspects of ToM, that is, it was less concerned with the composition of ToM in terms of underlying components (e.g. first order and second order beliefs).

In order to study the quantitative changes in ToM development, we opted for a cross-sectional design: children were tested only once and children of different ages were compared with one another. Such a design focuses on ’first encounter’-effects, i.e. how children react to explicit and formalized ToM-questions the first time they are confronted with it by means of a test. This procedure may provide a solid base to identify the potential critical points in the dynamics of ToM development and may result in a more profoundly documented starting-point for future time-serial

research. For, little to nothing is known about at which ages transition phenomena occur - such as accelerations, delays and regressions - in ToM development.

Since transition phenomena are non-linear, we applied similar non-linear smoothing techniques as in the norming procedure, namely Loess. Next to that, dynamic growth model building and additional statistical indicators – such as moving skewness, growth rate and variability – were used in order to demonstrate possible developmental transitions.

The findings showed that there is a three-step developmental model of basic ToM. The ToM sum-scores showed an overall developmental trend that leveled off towards the age of ten years. The greatest increase was obvious between 3.5 and 4.7 years of age. Within the overall increasing trend two statistically significant discontinuities were found, one at the age of 56 months and another at the age of 72-78 months (see Figure 1 in Chapter 4, pp. 79). The most pronounced discontinuity, the one at 72-78 months, was researched more into detail. It was accompanied by a decrease in growth rate and variability, and a change in skewness of the ToM data, all suggesting a developmental shift in ToM understanding. The statistically significant temporal decrease also occurred in the different ToM sub-scores and most clearly so in the core ToM component of beliefs (see Figure 6 in Chapter 4, pp. 88). The temporal decrease occurred in boys and girls alike. The discontinuity could not be accredited to gender differences, although it should be noted that girls had an earlier growth spurt than boys (see Figure 5 & 7 in Chapter 4, pp. 87 & 90 respectively). In Chapter four, the consequences of these findings are discussed from various theoretical points of view, with an emphasis on a dynamic systems interpretation of the underlying developmental paths.

With the achievement of the third objective of this dissertation, a new discovery was made, namely that there are non-linear phenomena in the ToM development of typically developing children. They evidence temporal regressions in their ToM development. Taking into account these temporal regressions is important to get more insight into the dynamics of ToM development. Next to that, the eventual absence or delay of these temporal regressions in children with ToM problems, may serve as an extra parameter in comparing children with and without ToM problems.

**ToM development in children with PDD-NOS**

Chapter five looks into the development of ToM in children with PDD-NOS. In doing so, it complies with the fourth objective of this dissertation.

As mentioned before, children with PDD-NOS are often not included in ToM research. However, they may be considered a distinct group concerning ToM problems. We hypothesized that children with PDD-NOS experience less severe ToM problems than autistic children, obvious mainly in quantitative differences but not in qualitative differences, indicative for a delayed ToM development instead of a deviant one.

Based upon quantitative and qualitative ToM aspects a two-dimensional ToM continuum was hypothesized (for a definition of quantitative and qualitative ToM aspects, see this Chapter, pp. 137, second paragraph). This continuum contains four quadrants (see Figure 1, in Chapter 5, pp. 102). It was hypothesized that the degree of ToM functioning is correlated to the severity of the autism diagnosis: children with PDD-NOS, typically developing children, children with an autistic disorder, and children with Asperger's syndrome occupy different quadrants. The study was aimed at examining the characteristics of the ToM problems and the ToM development in children with PDD-NOS.

In order to map ToM development a longitudinal research was undertaken. A group of 30 children with PDD-NOS was followed over a period of 20 months, and were tested every four months. This time interval is long enough to overcome trivial learning effects and short enough to follow ToM development closely. In addition, we made use of parallel versions of the ToM Storybooks version Sam (namely version Lotje, Pieter and Hanna) to further avert trivial learning effects due to mere repetition of the same test. Based upon the two-dimensional ToM continuum, it was hypothesized that the ToM development of children with PDD-NOS would be delayed, not deviant, in comparison to typically developing children. Next to that, it was hypothesized that children with PDD-NOS would probably also show a temporary regression in their development, as was found in typically developing children (see Chapter 4), but that this temporal regression would also be delayed.

Five questions were addressed: 1) Do children with PDD-NOS evidence ToM problems? 2) Are their ToM problems subject to age changes? 3) Do children with PDD-NOS evidence a developmental progress in ToM functioning? 4) Is the nature of this developmental progress subject to age changes? 5) Is their ToM growth also visible in their daily social behavior?

All questions were answered affirmatively. Children with PDD-NOS displayed ToM problems, but also a spontaneous ToM development. This development was delayed as was hypothesized. The same developmental sequence was followed but at a slower pace. Also the anomaly as known from the typically developing group was delayed in this group (an anomaly at the age of 85-90 months as compared to 72-78 months, see Figure 3 in Chapter 5, pp. 118). This anomaly was supported by an additional temporary increase in the coefficient of variation, indicative of a true discontinuous change. Taking into account the different timing of the anomaly in typically developing children versus children with PDD-NOS, the latter group of children can wrongly appear to be better or worse than their typically developing age mates. This is an important finding which should be taken into account when assessing the ToM of children at the age of 70-90 months old. The fact that the anomaly was also found in children with PDD-NOS makes the findings from Chapter four even more robust. It also provides additional evidence for a delayed ToM development in children with PDD-NOS. Next to the fact that they showed the same qualitative ToM score profile as typically developing children (see Figure 2 in Chapter 5, pp. 117), they also evidenced similar developmental phenomena.

Despite the precautions taken to prevent trivial learning effects due to simple recognition of items for instance, the ToM development of children with PDD-NOS was influenced by the repeated measurements. A classical learning curve was found (see Figure 5 in Chapter 5, pp. 121), consistent with findings from intelligence research and ToM research. The consequences of this finding for ToM trainings are discussed. On the one hand ToM training studies should correct their results for natural ToM development (for instance, one experimental group undergoes the ToM training, while the other - age and gender matched - experimental group undergoes a waiting list condition of the same period of time). On the other hand, training studies should also look into the mere effect of repeated testing (for instance, one experimental group undergoes the ToM training, while the other - age and gender matched - experimental group undergoes a repeated measurements design). From a learning theoretical point of view, repeated testing with long enough intervals and with parallel tests probably amounts to a form of exposure learning, i.e. exposure to items invites children to reflect on the questions and actively seek for answers.

Concerning the fifth question in the longitudinal study, it was found that ToM functioning was related to daily social skills. However, the

amount of ToM growth was not reflected in the VABS interview measuring the daily social skills. It was hypothesized that this might be due to the fact that the ToM of children with PDD-NOS is more cognitively mediated and less intuitively reflected than that of typically developing children. This difference may explain the weak link with daily social skills in children with PDD-NOS. Otherwise, also other autism related problems may have played a role, like executive function problems, a weak central coherence and generalization problems. Further research looking into this subject is needed.

Indirectly, this longitudinal research provided evidence for the hypothesized ToM quadrants, at least for one quadrant, which consists of a group that has been overlooked in previous research. It concerns a group of children with only quantitative ToM problems, not qualitative ones. This group has just a delay in ToM and not a deviant ToM. It is a group with PDD-NOS and not with an autistic disorder. Further longitudinal research is required to validate the other quadrants and to look into the nature of ToM problems in children with an autistic disorder and children with Asperger´s syndrome in comparison with children with PDD-NOS. Possibly neuro-imaging research could give further clues in disentangling the four quadrants. For instance, are brain regions differently involved when answering ToM tasks in individuals with Asperger's syndrome in comparison with individuals with PDD-NOS and neuro-typical individuals?

**Past and future of the ToM storybooks**

The ultimate goal of this dissertation was to get insight in the ToM development of children with PDD-NOS. However in order to reach this goal, three additional objectives were formulated, which were all attained. First, we constructed a comprehensive ToM instrument: the ToM Storybooks. This instrument incorporates different relevant basic ToM tasks and is applicable for a wide age range. Second, the ToM Storybooks proved to have good psychometric qualities and norm scores were calculated. Third, the development of ToM in typically developing children was described and showed a non-linear growth pattern, including a dip in ToM development. This dip proved to be an indicator for developmental discontinuity, not only in typically developing children, but also in children with PDD-NOS.

Recurring themes over the different chapters were gender differences in ToM functioning, quantitative versus qualitative differences in ToM functioning, developmental discontinuities in ToM functioning, and the use

of non-parametric techniques (Loess smoothing)and random permutation techniques (Monte Carlo analysis). They resulted in a host of new findings, in comparison with other ToM studies. These new findings concern gender differences (girls are more rapid in their ToM development), the occurrence of a dip in the development of ToM in typically developing children (serving as an indicator for discontinuity in ToM development), the delay of this dip in children with PDD-NOS, the developmental progress of children with PDD-NOS due to repeated testing, and filling one of the quadrants of the hypothesized two-dimensional ToM continuum.

Because psychometric qualities are good and norm scores are available, the ToM Storybooks can be used in applied and clinical settings, which on a small scale is already happening. The test is used at the University Centre for Child and Adolescent Psychiatry Groningen (a subdivision of Accare) and at the Autism Team North Netherlands (a subdivision of Lentis). In addition, the instrument has been and is used in several other research projects: a small study on discriminant validity of the ToM Storybooks comparing the ToM problems of children with PDD-NOS and children with ADHD (Blijd-Hoogewys et al., 2002), a small study on divergent validity concerning executive function problems, motor problems and sensory problems in children with ASD (Meindertsma, 2007), a small validation study in Chinese children (Teng, 2003), a small validation study of the ToM Storybooks in Finland (Vesterinen, 2008), a large replication study in Italy (Molina & Bulgarelli, 2007, in press; Molina et al., 2008), , a longitudinal study in Switzerland (Thommen), an effect study on supporting social and emotional development in young children (4 – 6 years old) through picture books (Kwant, ongoing), an effect study of a ToM training study (van den Broek, 2005), an effect study of cognitive behavior therapy in sexually abused children and children exposed to intra-familiar violence (Douma-Slothouber et al., submitted), an explorative study in hearing children from death parents (van Bavel, 2006), an explorative study in children who experience auditory hallucinations (Bartels, ongoing), an explorative study in children with a semantic language impairment (van Dijk, 2005; van Pol, 2005), an explorative study in children with an autism spectrum disorder and co-morbid mental retardation (Gordts & Nys, in preparation), a large replication study in developmentally disabled adult offenders (Solodova, ongoing), and a study of young children who undergo epilepsy surgery (Braams, ongoing).

## Strengths, limitations and recommendations

The ToM Storybooks provide a comprehensive, valid and reliable instrument for researchers and clinicians who wish to measure ToM in young children. Because there are norm scores up to the age of twelve, the test is particularly appropriate for comparison with clinical groups in which ToM development is delayed. Because of the use of simple ToM tasks and a motivating storyline, the test might also be useful in the field of intellectual disability, where autism spectrum disorders and related ToM problems are common.

Some limitations should be addressed. The test did not incorporate more complicated ToM tasks, such as second-order beliefs (inferring about what another person thinks about yet another person's thoughts; for a classic example see Baron-Cohen, 1989b). Also, there are fewer children tested in the older age range (from 8 years on), which implies a reduction in reliability at the older ages. Future research should include more complex ToM tasks and older age groups. We expect that this may lead to the discovery of additional anomalies in the development of ToM and that this may lead to an even better detection of children with subtle ToM problems.

Concerning future ToM research, we would like to stress the importance of explicitly searching for temporal discontinuities in ToM development. Next to that, ToM research should also consider dynamic testing principles - as opposed to static testing - where the learning potential of a child is quantified on the basis of his or her understanding and use of feedback given during testing (Grigorenko & Sternberg, 1998). This suggestion is supported by the finding that short term learning effects due to repeated testing (retested after 2 to 3 weeks) were only visible in typically developing children and not in children with PDD-NOS (see Chapter 2), whereas long term learning effects (retested every 4 months for a total period of 20 months) were visible in children with PDD-NOS. Including dynamic indexes in research can represent a quality step-up compared with static indexes (Fabio, 2005). Thus, constructing a dynamic testing version of the ToM Storybooks is highly advisable.

## Concluding remarks

There are three main highlights in this dissertation. First, we developed a comprehensive ToM test, consisting of a wide spectrum of basic ToM tasks and applicable to a wide age range. The test is valid, reliable and normed.

Second, typically developing children were found to evidence a non-linear development of their ToM, at least when based upon 'first encounter' measurements. They showed a three-step developmental increase. Two temporal discontinuities were found, one at the age of 56 months and another at the age of 72-78 months. This is an important empirical finding, which has not been reported on before. The thorough analyses of these discontinuities showed that these temporal discontinuities are genuine and not a statistical artifact of sampling flaws.

Third, children with PDD-NOS were found to have a delay in their ToM development. They also evidenced a temporal discontinuity, which was comparably delayed. In addition, they showed a learning effect due to repeated exposure to similar yet differently formulated test items (parallel versions of the ToM Storybooks).

As a result, we believe this dissertation is innovative for different disciplines. Firstly, for developmental research, its contribution lies in the use of non-linear smoothing techniques, which have great value in researching development and which enable the researcher to identify anomalies in development, such as temporary regressions. Secondly, our study contributes to norming research by introducing a different way of calculating norm scores - based upon a non-linear smoothing method - and introduces a new method to calculate more valid confidence intervals. Thirdly, its relevance for autism research is that it considered children with PDD-NOS as a major research group. And finally, it contributes to ToM research in that it has shown temporary discontinuities in ToM development, a two-dimensional ToM continuum, and a delayed ToM development in children with PDD-NOS.