Habituation of Premonitory Urges as Working Mechanism of Exposure and Response Prevention in the Treatment of Tourette’s Syndrome and Tic Disorders

Nelleke van den Berg
340914
Klinische Kind- en Jeugd Psychologie
Erasmus Universiteit Rotterdam
2013-2015
**Introduction**

The Gilles de la Tourette Syndrome (GTS) or Tourette Syndrome (TS), first described by George Gilles de la Tourette in 1885 as ‘la maladie des tics’ (Tourette, 1885), is a neurological disorder characterized by multiple, simple or complex, motor and one or more vocal tics (APA, 2013). Besides the presence of motor and vocal tics, the diagnosis for this tic disorder is based on the duration of the tics (they can vary in frequency, but have persisted for more than 1 year since first tic onset), the onset (before the age of 18 years) and absence of other possible causes, like medical conditions or substance use.

It is complex to determine the prevalence of Tourette Syndrome, partly because of methodological problems like the complexity of the diagnosis and the high rate of comorbidity in Tourette. Based on several international studies, Tourette Syndrome has an overall international prevalence of 1% in youngsters aged 5-18 years (with a range from 0.4% to 3.8%) (Robertson, 2008; 2011, 2012). The age of onset of TS symptoms ranges from 2 to 21 years (mean age 6-7 years), usually marked by the occurrence of simple, motor tics (Cavanna, Servo, Monaco & Robertson, 2009; Robertson, 2002). Males have more commonly TS than females, with male-female ratio varying from 2:1 to 4:1 (APA, 2013). Although there are gender differences in prevalence, there are no gender differences in the kind of tics, age at onset or course.

In TS, the tics are often preceded by an aversive or unpleasant sensation of a somatosensory nature, a so-called ‘premonitory urge’ or ‘premonitory sensation’ (Bliss, 1980; Cavanna & Rickards, 2013; Kwak, Vuong & Jankovic, 2003; Leckman, Walker & Cohen, 1993; Woods, Piacentini, Himle & Chang, 2005). This unpleasant sensation can manifest itself for example in the skin, muscles, joints, bones or vocal cords and involves the experience of, for example, pressure, energy, prickling, pain or itching. After the manifestation of the tic, there is a temporary relief or reduction of the sensation. The premonitory sensations can be considered as the involuntary component of the tics, while the tics, which are executed in response to the unpleasant premonitory sensation, have a more intentional character (Bliss, 1980; Cavanna & Rickards, 2013; Kwak, Vuong & Jankovic, 2003).

TS is in approximately 90% associated with comorbid psychopathology and behaviour disturbances (Cavanna et al., 2013c; Cavanna & Rickards, 2013; Robertson, 2002, 2012).
The most common comorbid disorders in TS are attention deficit hyperactivity disorder (ADHD), obsessive compulsive disorder (OCD), autistic spectrum disorders (ASD) and affective disorders like depression. Furthermore, TS is associated with antisocial behaviours, oppositional defiant disorder (ODD), conduct disorder (CD), learning difficulties, anxiety and personality disorders (Cavanna & Rickards, 2013; Cavanna, Servo, Monaco & Robertson, 2009; Robertson, 2002, 2012).

Quality of life

TS has a significant impact at the health related quality of life of individuals with Tourette Syndrome (Cavanna et al., 2008). ‘Health Related Quality Of Life’ (HR-QOL) can be defined as ‘the perceived physical, mental and social effects of an illness and associated therapies on a patient over time’ (World Health Organization; Cavanna et al., 2013).

Several studies suggest that children and adolescents who experience tics, have lower quality of life than healthy individuals (Colonea et al. 2011; Eddy et al., 2011; Elstner, Selai, Trimble & Robertson, 2001; Jalenques et al., 2012; Muller-Vahl et al., 2010; Storch et al., 2007). Conelea et al. (2011) assessed the tic-related functional impairment in a sample of youth with chronic tic disorder (CTD) and their parents. Based on their findings, they suggest that this youth experience mild to moderate adverse impact on various domains of functioning. Higher tic severity is associated with greater functional impairments in various domains and affects of life (Colonea et al., 2011; Cutler, Murphy, Gilmour & Heyman, 2009; Storch et al., 2007), except physical functioning (Muller-Vahl et al., 2010; Storch et al., 2007). These findings indicate that the presence of tics has a negative impact on patients lives. Both children and adults are most impaired in emotional and social domain (Cavanna, 2013a). Children and youth with TS experience interference with social functioning, friendships, social life and family relations. Furthermore they experience social difficulties like bullying, teasing, stigmatization and tic-related discrimination, as well as people in their environment not understanding the involuntary behaviour and lack of control of the tics. This is associated with social withdrawal, the avoidance of public places and group activities and family stress (Colonea et al., 2011; Colenea et al., 2013; Cutler et al., 2009; Eddy et al., 2009). In adulthood TS can also cause social inference, as well as psychological problems, like (social) anxiety, depression, social phobia, panic and agoraphobia. Also occupational interference is reported, for example a negative impact of tics at work productivity and being late due to tics. Finally, TS can lead to physical inference when tics cause pain of physical damage (Colenea et al.,
This significant impact of TS at the health related quality of life of individuals with Tourette Syndrome, emphasizes the importance of good treatment of tics, which can contribute to a better quality of life.

Treatment

Because of the negative impact of tics on various domains of functioning, it is important to provide a good treatment. The treatment of TS is complex and a multidisciplinary treatment is recommended (Hartmann & Worbe, 2013; Robertson, 2012). First of all, it is important to provide psychoeducation to the patients and their families as well as to the school or work environment (Hartmann & Worbe, 2013; Robertson, 2012; Verdellen, van de Griendt, Hartmann, Murphy, 2011). Providing information about the nature of tics, co-morbidities and overall prognosis of the disorder promotes understanding and acceptance of TS and the condition of the patient, and it reduces misunderstanding, uncertainty, stress and stigma in TS (Roessner et al., 2013; Verdellen et al., 2011). Based on the nature and complexity of the tics, comorbidity, subjective discomfort, quality of life and the impairment of social, emotional, physical and school- or work functioning, an appropriate intervention can be provided (Hartmann & Worbe, 2013; Roessner et al., 2013).

Pharmacotherapy can be an effective treatment for moderate to severe tics, comorbidity and psychopathology in TS patients. Dopamine antagonists, typical antipsychotics or neuroleptics, are important in the pharmacological treatment of TS, because they affect the altered dopamine transmission in TS (Hartmann & Worbe, 2013; Scahill et al., 2006; Thomas & Cavanna, 2013). Well-established, effective medications for the treatment of tics are the neuroleptics haloperidol and primozide, the atypical antipsychotic risperidone and clonidine (Hartmann & Worbe, 2013; Peterson & Azrin, 1993; Robertson, 2011; Scahill et al., 2006; Thomas & Cavanna, 2013). In case of potentially dangerous, isolated or important vocal tics, especially for those in the neck muscles or vocal cords, a botulinum toxin injection is an effective treatment (Hartmann & Worbe, 2013; Kwak, Hanna & Jankovic, 2000; Scahill et al., 2006).

Verdellen, van de Griendt, Hartmann and Murphy, working group of the European Society for the Study of Tourette Syndrome (ESSTS; 2011), have published a clinical guideline with recommendations for the behavioural and psychosocial interventions (BPI) of children and adolescents with tic disorders, based on a systematic literature search to the efficacy and effectiveness of BPI for tic disorders. Several effective behavioural treatments
for tics, described in this clinical guideline, are massed (negative) practice (MP), habit reversal training (HRT), self-monitoring (SM), contingency management (CM)/function based interventions (FBI), relaxation training (RT)/hypnosis, exposure and response prevention (ERP), cognitive behavioural treatment (CBT) and bio(neuro)feedback (NF).

Habit reversal and exposure and response prevention

The literature provides support for habit reversal training and exposure and response prevention as evidence-based, equally effective treatments for tics for both children and adults (Verdellen et al., 2011). Habit reversal (HR) is an intervention for eliminating nervous habits and tics, developed by Azrin and Nunn (1973), and is currently the most widely used strategy for the treatment of tics (Verdellen et al., 2008; Verdellen, Hoogduin & Keijsers, 2007). The rationale of this treatment is that tics and nervous habits are learned responses or adapted startle reflexes. According to Azrin and Nunn (1973), the nervous habit starts as a normal reaction or reflex to a traumatic event or physical injury. When this reaction persists after the traumatic event and adopts an unusual form and unusually high frequency, it is called a nervous habit. This habit has an automatic nature and is performed so often that it evades personal awareness. So an originally normal reflex develops into a tic. The tic persists because of response chaining, limited awareness, excessive practice, social reinforcement and tolerance. To counteract the reinforcing factors which maintain the tics, in HR patients learn to become aware of the tics through response description, response detection, early warning and awareness training. To inhibit the tic, they practice a competing response of the tic with competing response training. This response is a movement that makes it impossible for the tic to come through. To maintain the inhibition of the tic and performing of the competing response, the patients receive motivation enhancement and generalization training (Azrin & Nunn, 1973; Griendt, Verdellen, Dijk & Verbraak, 2013). Both controlled and uncontrolled studies indicate that habit reversal is an effective intervention for tics, with tic reductions of 50% to 98% (for a review, Carr & Cong, 2005; Cook & Blacher, 2007; Peterson & Azrin, 1993; Verdellen et al., 2011; Woods & Miltenberger, 1995).

Exposure and response prevention (ERP) is a treatment based on the ideas of Bliss (1980) that motor and vocal tics are produced voluntarily to relieve the unpleasant premonitory sensation (Bliss, 1980; Hoogduin, Verdellen & Cath, 1997; Leckman, Walker & Cohen, 1993). The association of the sensation followed by a tic that relieves the sensation, will be strengthened over time. The aim of ERP is to interrupt this association and prevent the
tics to occur. So in ERP the individuals are exposed to the sensations and urges that precede the occurrence of the tics, followed by the response prevention of the tics. While resisting all tics (response prevention), the patient habituates to the premonitory sensations (so-called ‘habituation’), what results in tic reduction (Griendt et al., 2013; Hoogduin, Verdellen & Cath, 1997; Verdellen et al., 2004; Verdellen et al., 2008). This habituation process is depicted in Figure 1 and can be described like this: a person experiences an unpleasant sensation, the ‘premonitory sensation’, preceding a tic. When the tic manifest itself, there is a relief or reduction of the premonitory sensation. But this relief or reduction is temporary and the sensation will increase again, followed by a tic. In exposure and response prevention treatment the person learns to prevent the tics to occur, meanwhile exposed to the increasing unpleasant sensation. (S)he learns and experiences that the unpleasant sensation will reach a top, followed by a decrease. The more the person trains with resisting the tics and undergoes the flow of the premonitory sensation, (s)he will habituate to this process and the top of the sensation-curve will decrease. Finally, the person can endure the notable increased sensation without continuously manifesting tics.

A pilot study of four patients with a severe form of TS suggest that exposure and response prevention treatment can decrease tic behaviour (Hoogduin, Verdellen & Cath, 1997), with a tic reduction at the end of treatment of 68% to 83%, and an tic reduction at follow up two months later of 50% to 100%. A randomized controlled trial in 43 patients (Verdellen et al., 2004) has shown the efficacy of exposure and response prevention in reducing tic symptoms. After treatment with exposure and response prevention there were clinically significant improvements of tic behaviour in 42% of the patients.

Habituation

A few studies examined the habituation to the premonitory sensations during the exposure and response prevention treatment (Specht et al., 2013; Verdellen et al., 2008). Verdellen et al. (2008) tested the habituation hypothesis as possible underlying mechanism of ERP in 20 TS patients in 10 two-hour sessions. Before each session and at 15-minute intervals the severity of the tics was rated. Corresponding with reports of TS patients that sensory

Figure 1. Habituation curve
sensations and urges initially intensify with tic suppression (Fahn, 1993; Kurlan, Lichter & Hewitt, 1989), Verdellen et al. (2008) expected a quadratic habituation curve (a curve showing an increase, followed by a decrease, see Figure 2), with an intensification of sensory sensations and urges with tic suppression, followed by a decrease of severity of premonitory sensations.

This is in line with findings in the treatment of OCD with ERP (Marks, 1987; Rachman & Hodgson, 1980) and research of Walton, Gaind and Marks (1972), where continuous exposure to intense phobic stimuli initially evoked an intensification of anxiety reaction (risen heart rate), and is followed by a decline. Verdellen et al. (2008) found support for habituation as a mechanism of change in ERP in TS. Significant reductions of the severity of sensory sensations were found during the treatment sessions (within-session habituation) and a significant decrease in severity of sensory sensations across sessions (between-session habituation). In contrast to the expected quadratic trend in severity of sensations (an initial increase of the severity of the sensations, followed by a decrease over time), they found an linear reduction. A possible interpretation for this finding is that the 15-minutes interval was too long to detect initial increase of severity of sensory sensations.

Specht et al. (2013) examined the ability of youth to suppress tics for prolonged periods and the possible subsequent rebound in tic severity. They measured the premonitory urge severity at 10-seconds intervals in 12 patients (ages 10-17 years). Consistent with findings of Verdellen et al. (2008), there was no significant initial increase in premonitory urge ratings during tic suppression. Contrary to the findings of Verdellen et al. (2008), there was no habituation to the premonitory urge during prolonged tic suppression. Instead, the urge intensity ratings remained relatively stable. A possible explanation for the relatively stable urge intensity is that the 10-second-interval was too short to detect initial increase. By asking for the intensity of the urge every 10 seconds, the urge is possibly continuously triggered, and as a result there is no increase of decrease in intensity, but a stable pattern.
These different findings in research concerning habituation lead to the question: is there a quadratic habituation trend to premonitory urges as underlying mechanism in ERP?

Figure 3. Different habituation curves: quadratic trend, linear trend and a stable pattern

The aim of the present study is to test habituation as underlying mechanism in ERP treatment in TS by rating the premonitory urge severity in shorter intervals than Verdellen et al. (2008), and longer intervals than Specht et al. (2013). The main hypothesis is, in line with reports from patients, that there is a quadratic habituation trend to premonitory urges with an intensification of sensory sensations and urges with tic suppression, followed by a decrease of severity of premonitory sensations, during the treatment sessions (within-session habituation). To test the expected initial increase of premonitory urge severity, this urge severity is measured every 15 minutes, and in the first quarter every 5 minutes. The second hypothesis is that the perceived severity of premonitory sensations will decrease across sessions (between-session habituation).
**Method**

**Participants**

In this study, 29 patients (20 men, 9 women; mean (SD) age 17.97 (12.15) years; age range, 7 to 59 years), referred by neurologists, psychiatrists and general practitioner to two outpatient clinics specialized in the treatment of tic disorders, were asked for participation on a voluntary base. They gave written informed consent, and in case of children, their parents did as well. Inclusion criteria for this study was a diagnosis of Tourette Syndrome (TS) or Chronic Tic Disorder (CTD) in accordance with the Diagnostic and Statistical Manual of Mental Disorders (4th ed.; DSM-IV; American Psychiatric Association, 1994). Exclusion criteria were present comorbidity of major depression, psychotic disorder and mental retardation. The use of medication for TS and CTD was not an exclusion criterion, as long as daily doses were not increased nor type of medication was changed throughout the study. Inclusion and exclusion criteria were independently checked by two experienced clinicians.

**Material**

Several measurements at different times during the treatment have been performed. For adults, the Symptom Checklist -90 (SCL-90) and the ADP-IV (Assessment of DSM-IV Personality Disorders) were used as screening instruments for psychopathology. For children, the Child Behaviour Checklist (CBCL), the Teacher Report Form (TRF) and the Youth Self Report (YSR) were used. The results of this measurements are not of primary use in this study. To obtain information about the severity of tics, the Yale Global Tic Severity Scale (YGTSS) and the Premonitory Urge for Tics Scale (PUTS) were used.

The SCL-90 is a widely used multidimensional measure of psychological distress and somatic complaints in both clinical practice and research (Derogatis, 1975;1977). Arrindell and Ettema (1975, 1986, 2005) have developed a Dutch version of the SCL-90, the Symptom Checklist. Research shows that the SCL-90 is a reliable and valid assessment tool, with Cronbach’s Alpha ranging from $\alpha = .82$ to $\alpha = .92$ and reliability of the entire test of $r = .94$ (Arrindell & Ettema, 1981; Brophy, Norvell & Kiluk, 1988).

The ADP-IV is a self-report questionnaire for personality disorders defined by Axis II of the DSM-IV (APA, 1994; Schotte & Doncker, 1994; Schotte, Doncker, Dmitruk, Valck & Mulders, 2002). The ADP-IV has a good validity and reliability. The internal consistency of
the ADP-IV dimensional scales is adequate, with alpha coefficients ranging from $\alpha = .60$ to $\alpha = .88$ (Schotte & Doncker, 2000; Schotte, Doncker, Vankerckhoven, Vertommen & Cosyns, 1998). The interrater reliability is satisfactory as well, with Cohen’s Kappa $\kappa = .67$ (Schotte & Doncker, 2000).

The CBCL, YSR en TRF are part of the Achenbach System of Empirically Based Assessments (ASEBA). The CBCL is a parent-rated questionnaire, aimed to assess the frequency and intensity of behavioural, emotional and social problems of the child (Achenbach, 1991; Achenbach & Ruffle, 2000). The CBCL is a good, cross-cultural valid questionnaire (Masha et al., 2007), with a strong content validity, a good internal consistency with alpha $\alpha = .73$ to $\alpha = .79$ and a high test-retest reliability with a mean alpha $\alpha = .73$ to $\alpha = .79$ (Achenbach & Rescorla, 2000, 2001). The TRF (rated by the teacher) and the YSR (completed by the child or adolescent) also assess the emotional and behavioural functioning of children (Achenbach, 1991), and has a good internal consistency and test-retest reliability as well (Achenbach & Rescorla, 2000, 2001).

The primary outcome measure was the Dutch version of the Yale Global Tic Severity Scale. The YGTSS (Leckman et al., 1989) is a semistructured clinical-rated instrument to assess the nature, number, frequency, intensity, severity and complexity of the motor and phonic tics, based on child and parent reports and behavioural observations. The YGTSS consist of tic inventory (2 checklists) and ordinal scales (Total Motor Tic Score, Total Phonic Tic Score, Total Tic Score, Overall Impairment Rating, and Global Severity Score). The ordinal scales are rated on a six-point scale from 0 to 5. A global severity score is obtained by combining all ratings, and is rated on a 1 scale of 0 (none) to 50 (severe). The YGTSS can be used to obtain a baseline measure prior to the implementation of an intervention, as an indicator of change during and following an intervention (Leckman et al., 1989; Storch et al., 2005). The YGTSS has a good validity and reliability, with high ICC (.80 to .91) for the subscale scores (Leckman et al., 1989) and high Cronbach’s alpha reliability coefficients ranging from $\alpha = .92$ to $\alpha = .94$ for the subscale scores (Storch et al., 2005).

The PUTS is a brief self-report assessment instrument to measure premonitory urge in youths with CTD or TS (Woods, Piacentini, Himle & Chang, 2005). Children have to rate on a four-point scale the extent to which the tics are preceded by several urges. The PUTS is a reliable and valid assessment instrument of the premonitory urge phenomenon, with a good internal consistency ($\alpha = .81$) and a good test-retest reliability (one-week test-retest reliability
\( r = .79, p < .01; \) two-week test-retest reliability \( r = .86, p < .01; \) Woods, Piacentini, Himle & Chang, 2005).

To assess the presence of comorbidity, the Children's Yale-Brown Obsessive Compulsive Scale (CY-BOCS), the Conners test for ADHD and the Vragenlijst voor Inventarisatie van Sociaal gedrag van Kinderen (VISK; English version: Children's Social Behaviour Questionnaire, CSBQ) were used. These tests are outside the scope of this research.

Design and procedure

Treatment

Treatment consisted of two training sessions and ten actual ERP sessions, with a duration of one hour. In the training sessions, patients learned to systematically suppress tics for a prolonged period. If one tic in particular came through three times consecutively, time was stopped and attention was focused at that specific tic, until the patient could suppress it for at least 5 minutes. At 0, 30 and 60 minutes, the severity of the urge was rated at a five-point Likert scale (0-4; the so-called SUD-score), for at most seven different tics. The therapist functioned as a coach, encouraging the patient to improve his/her achievements. In the following 10 ERP sessions, patients were asked to suppress their tics for 1 or 2 (consecutive) hours, with the addition that they had to focus on the premonitory urges. To optimize exposure, patients were asked to concentrate on the premonitory urges in their body, the therapist talked about the urges, and urge-eliciting objects were introduced in the session (e.g. exciting games, videotapes of tic programs etc.). At 0, 5, 10, 15, 30 and 60 minutes, urges were rated at a five-point Likert scale (0-4; SUD-score) again for at most seven different tics. Regardless of the intensity of urges, tics needed to be suppressed all the time. The therapist motivated the patients to do their ultimate best. Patients were told they could profit from treatment the most if they practiced exercises at home as well. To enhance treatment integrity, treatment manuals were used.

Procedure

The total procedure consist of treatment and measurement moments. At the intake session, the SCL-90 and ADP-4 were administered in adults and the CBCL, TRF and YSR were administered in children and youth and their parents/teachers. In case of comorbidity, the CY-BOCS, Conners and VIS-K were administered. This measurements are not used in
this study. Before the first training session (M0), the severity of the tics were measured with the YGTSS, PUTS and Tic Inventarisatielijst, this measurements are of primary use in this study. Furthermore, use of medication was registered and the tics were taped on video for 15 minutes. To have a good representation of the tics, a parent or partner of the patient has to register his/her tics in daily life for 15 minutes a day. In the next two sessions, the training sessions, patients learned to systematically suppress tics for a prolonged period. Between this training sessions and the first ERP session, tic severity was measured with the YGTSS and the PUTS, and the tics were videotaped for 15 minutes (M1).

To measure any changes in the perceived severity of the premonitory sensations, before treatment, for each patient one to seven individual sensations corresponding to the most frequent or interfering tics at that time, were selected. Global tension was added as an item for each patient. Before each ERP session and at 5, 10, 15, 30, 45 and 60 minutes, therapists asked the patients to rate the severity of these maximal eight items on a Subjective Units of Distress Scale ranging from 0 (absent) to 4 (very severe). Thus, a total sensation severity score (SUD-score) was obtained (range = 0–32, depending on how many sensations were rated).

In addition, therapists registered the occurrence of each tic (i.e., not only the seven tics selected for the measurement) that had not been suppressed.

After five ERP sessions the tic severity was measured again with the YGTSS and the PUTS, and the tics were videotaped again (M2). After the last ERP session the tic severity was measured with the YGTSS and the PUTS, and the tics were videotaped for 15 minutes (M3). Furthermore was the diagnosis was checked at end-of-treatment (if patient still met the DSM-IV criteria for TS). Finally, a follow-up measure (M4; similar as M3) was performed three months after the ERP treatment.
Statistical Analysis

To describe the study population standard descriptive statistics were used in SPSS (IBM Corp., 2013; version 22.0). The YGTSS scores were assessed by the therapist and the videos were scored by independent assessors. The inter-observer agreement between the therapists and the assessor was calculated with a Pearson correlation coefficient. Fifteen of the 29 videotapes (51.7%) were scored by two independent assessors and the inter-observer agreement between the two independent assessors was calculated with a Pearson correlation coefficient.

Main hypothesis within session-habituation

To test the main hypothesis, that there is a significant reduction in severity of premonitory urge (SUD-score) within the sessions (within-session habituation), an average sensation severity model will be used. The average sensation severity score over patients over sessions (for the global tension item and maximal seven selected tics) is calculated in this model (average within SUD-score).

A steps-wise procedure of statistical analyses is performed, to achieve the best model fit to the data. First, a Repeated Measures ANOVA (RMA; using SPSS) is used to test if there is a significant decline in this average SUD-score. Furthermore a quadratic regression function is used to describe the development of SUD-scores during the treatment sessions and to evaluate if there is a quadratic habituation trend.

There are several properties of SPSS that make a less model fit with our data (Field, 2009). The study population consists of 29 participants. This is a relative small sample size, which makes it hard to attain statistical significant or meaningful results as well as sufficient coverage and power.

Besides, SPSS procedures perform listwise deletion of records: a case is excluded from the analysis because it has a missing value in at least one of the specified variables. The analysis is only run on cases which have a complete set of data. In our dataset are at least 70 measurement moments and several cases have one or more missing data. In consequence, many cases will be excluded when performing a RMA. Furthermore, SPSS does not take into account the individual differences of the participants. Every case is forced to follow the same trajectory, by fixing one intercept and one slope. Finally, in SPSS the assumption is that every time interval between the several measurements is the same. In our dataset the interval

13
between the first four measurement moments (0, 5, 10, 15) is a five-minute interval and the
interval between following measurement moments (15, 30, 45, 60) is a fifteen-minute
interval. In this case, using SPSS will not result in an optimal fit with the data.

To prevent the listwise deletion in SPSS, Confirmative Factor Analyses (CFA) are
executed in Mplus (version 6.12, Muthen & Muthen, copyright 1998-2011). CFA attempts to
determine the minimum number of interpretable factors (incl. continuous latent variables (i.e.
missing values, measurement errors)) that can adequately describe the correlations among a
set of observed variables. This model does not impose a structure on the linear relationship
between the observed variables and the continuous latent variables, it only specifies the
number of latent variables (Mplus, retrieved from www.statmodel.com). Using CFA, factor-
scores of the average SUD-score over patients over sessions (for the global tension item and
seven selected tics) are computed.

Finally, a Latent Growth Model (LGM), executed with the program Mplus (version
6.12, Muthen & Muthen, copyright 1998-2011) using time trends, is used to test whether a
quadratic habituation curve could be found in the data. The factor-scores, computed with
CFA, are inserted in Mplus. The Latent Growth Model uses a quadratic, a linear and a piece-
wise regression function to describe the relationship between time and SUD-score for each
combination of patient and session number. Advantages of LGM (relative to RMA in SPSS)
are that there is no listwise deletion, individually-varying times of observation are read as data
and multilevel models cope with random effects. Furthermore, LGM takes into account the
individual differences of the participants and every cases has a unique trajectory with an
intercept and slope and an individually development over time (Tabachnick & Fidell, 2012).
Finally, with Mplus the model fit can be computed.

Second hypothesis between session-habituation

The second hypothesis is that there is a significant decrease in severity of premonitory
sensations (SUD-score) across sessions (between-session habituation). For this purpose, the
average sensation severity score over patients for each session is calculated (average between
SUD-score). To test if the average sensation severity model of the development of this
average SUD-score across sessions shows a significant decline, a Repeated Measures
ANOVA is used.
Results

Descriptives

The sample exist of 29 patients (20 male, 9 female; mean (SD) age 17.97 (12.15) years; age range 7 to 59 years), diagnosed with Tourette Syndrome (TS; \(N = 26; 89.6\%\)) or Chronic Tic Disorders (CTD; \(N = 3; 10.4\%\)). Twenty-one patients (65.5\%) were aged 18 years or younger. Mean age at tic onset of the sample was 6.59 years (\(SD = 2.49;\) range 3-13) and the mean duration of the TS symptoms was 9.07 years (\(SD = 8.13;\) range 1-36). The vast majority of the participants had no comorbid disorder \(N = 25; 86.2\%\), three persons had comorbid disorders, respectively ADHD, ADHD & Intermittend Explosive Disorder (IED), and Generalized Anxiety Disorder. Eleven patients used medication (37.9\%), respectively Risperidone, Pipamperon, Pimozide, Atomoxetine, Olanzapine, Aripiprazole, Clonidine, Moclobemide and Tetrabenazine. Patients and parents were told not to change the medication or enlarge the dose. The patients participated in 1 hour exposure and response prevention (ERP) sessions at the HSK Expertise Tic Centre. Data were collected between May 2008 and May 2009.

*Table 1. Characteristics of the patient sample (N = 29)*

<table>
<thead>
<tr>
<th>Frequency (%)</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sexe</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>20 (69%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>9 (31%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>17.97</td>
<td>12.15</td>
<td>7</td>
<td>59</td>
</tr>
<tr>
<td><strong>Diagnosis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TS</td>
<td>26 (89.6%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTD</td>
<td>3 (10.4%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Onset</strong></td>
<td>6.59</td>
<td>2.48</td>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td><strong>Duration</strong></td>
<td>9.07</td>
<td>8.13</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td><strong>Comorbidity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>25 (86.2%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD</td>
<td>1 (3.4%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADHD and IED</td>
<td>1 (3.4%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAS</td>
<td>2 (6.9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Medication</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>18 (62.1%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>11 (37.9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Interobserver Agreement YGTSS Scores

To analyze the interobserver agreement on the YGTSS score between the therapists and the assessors, a Pearson correlation coefficient was calculated. A high level of agreement was obtained for the YGTSS score measured before the treatment (M0; \( r = .70; p < .01 \)) and for the YGTSS score measured after the treatment (M3; \( r = .78; p < .01 \)). A Paired-Samples T-test showed that at M0 the difference between the therapists YGTSS score (\( M = 21.31, SD = 9.76 \)) and the assessors YGTSS score (\( M = 20.19, SD = 8.13 \)) was not significant, \( t(25) = .81, p > .05 \). The difference between the therapists YGTSS score (\( M = 13.83, SD = 7.96 \)) and the assessors YGTSS score (\( M = 13.42, SD = 6.63 \)) at M3 was also not significant, \( t(23) = .41, p > .05 \). Hence, the difference between the therapists and assessors scores is not significant; the therapists and assessors measured equally.

The interobserver agreement between two independent assessors was also calculated with a Pearson correlation coefficient. A high level of agreement was obtained for the YGTSS score measured before the treatment (M0; \( r = .83; p < .05 \)) and for the YGTSS score measured after the treatment (M3; \( r = .92; p < .05 \)). A Paired-Samples T-test showed that at M0 the difference between the YGTSS score of assessor 1 (\( M = 21.00, SD = 6.10 \)) and the YGTSS score of assessor 2 (\( M = 18.75, SD = 7.46 \)) was not significant, \( t(7) = 1.54, p > .05 \). The difference between the YGTSS score of assessor 1 (\( M = 11.60, SD = 7.70 \)) and the YGTSS score of assessor 2 (\( M = 11.80, SD = 7.40 \)) at M3 was also not significant, \( t(4) = - .15, p > .05 \).

Scores at PUTS and YGTSS

The mean score at the Premonitory Urge for Tics Scale (PUTS) measured before the treatment (M0) was 23.3 (\( SD = 5.66; \) range 10-34; see Appendix Table 2). During the treatment there was a small increase in mean PUTS score, with a mean score of 23.59 (\( SD = 5.61; \) range 13-35) after two treatment sessions (M1) and a mean score of 24.13 (\( SD = 5.01; \) range 13-32) after seven treatment sessions (M2). At the end of the treatment (after twelve sessions) the mean PUTS score was 23.52 (\( SD = 5.17; \) range 13-32; Figure 4). With a Paired-Samples T-test is investigated if there is a significant difference between the pre- and post PUTS score. As mentioned earlier, there was a small increase in mean score, but this increase was not significant, \( t(24) = -1.18, p > .05 \). The follow-up measurement of PUTS score was completed by 13 persons, with a mean score of 22.69 (\( SD = 6.32; \) range 12-32). According a
Paired Samples T-Test, the difference between pre-score and follow-up score is not significant, $t\ (12) = 0.14, p > .05$.

Figure 4. PUTS score development over time.

The mean score at the YGTSS before the treatment sessions (M0) was 18.81 ($SD = 7.27$; range 3-34.5). At the end of the treatment (M3) the mean YGTSS score was 12.54 ($SD = 6.8$; range 0-29.5). With a Paired Samples T-Test (see Appendix Table 3 and 4) is investigated if there is a significant difference between this pre- and post YGTSS score. On average, the patients had a significant higher YGTSS score before treatment ($M = 18.81, SD = 7.27$) than at the end of the treatment ($M = 12.54, SD = 6.80$), $t\ (25) = 3.80, p < .01$.

Assumptions and Outliers Main Hypothesis

The assumptions of the parametric data are tested, respectively the assumption of independent observations, normally distributed data and sphericity.

Independent observations: the scores in the experimental conditions are non-independent for a given participant, but behaviour between different participants is independent. This means this assumption is met.

Normally distributed data: to inspect the normality, the P-P plots (probability-probability plots), histograms of the standardised residuals and the values of skewness and kurtosis are checked. The data of the different measurement moments are normally distributed.

Sphericity: sphericity refers to the equality of variances of the differences between levels. With Mauchly’s test in SPSS the hypothesis that the variances of the differences between conditions are equal is tested. If Mauchly’s test statistic is significant ($p > .05$), there
are significant differences between the variances of the differences, and the condition of sphericity is not met. If Mauchly’s test statistic is not significant (\(p < .05\)), the condition of sphericity is met. In this dataset, Mauchly’s test indicated that the assumption of sphericity has been violated \(\chi^2 (20) = 186.37, p < .01\), therefore the Greenhouse-Geisser estimates of sphericity are reported (\(\varepsilon = .25\)).

Repeated Measures Analysis Within Session Habituation

The results of the Repeated Measures Analysis describe what happens within a session, averaged over patients and sessions. The results (see Appendix Table 5 and 6) show that the mean SUD-scores significant differ during sessions, \(F (1.49, 41.63) = 5.81, p < .05\), but there is not a reduction. Hence, the main hypothesis, that there is a significant reduction in severity of premonitory urge (SUD-score) within the sessions (within-session habituation), is not confirmed. The mean SUD-score during sessions is depicted in Figure 5, the development seems to follow a linear trend.

![Figure 5. Graphic representation of the within-session habituation. Mean sensation severity scores (SUD-scores) within sessions, averaged over patients (n = 29) and sessions (n = 10).](image)

According to the Test of Within-Subjects Contrasts (see Appendix Table 7), there is a significant linear effect, \(F (1, 28) = 5.50, p < .05\), and a significant quadratic effect, \(F (1, 28) = 14.71, p < .01\). The quadratic model has better fit with the data. This trend is visible when the scale of the Y-axis is adjusted from ‘0 to 4’ to ‘1.0 to 1.6’ (minimal mean = 1.11, maximal
mean = 1.47), see Figure 6. The hypothesis that there is a quadratic habituation trend to premonitory urges is confirmed in the RMA.

![Graph](image)

*Figure 6.* Close-up of the within-session habituation. Mean sensation severity scores (SUD-scores) within sessions, averaged over patients (n = 29) and sessions (n = 10).

**Latent Growth Model Within Session Habituation**

To test the hypothesis if a quadratic curve could be found in habituating to sensations, data were put into MPlus. First several Confirmative Factor Analyses have been performed, on six of the maximal seven sensations. It was not possible to take into account all seven sensations, because there were too many missing values on the 7th sensation. To filter out missing data in the first six sensations, Confirmative Factor Analyses were performed. The factor scores were put into a Latent Growth Model, resulting in the data in Table 8. Besides Linear and Quadratic curves, also a Piecewise model was tested, because of expected effects in the first 15 minutes. See Appendix for a graphic representation of the Linear, Quadratic and Piecewise curves (Figure 7).
Table 8. Best Fitting Model on the Data

<table>
<thead>
<tr>
<th>Model Type</th>
<th>AIC / BIC</th>
<th>CFI</th>
<th>TLI</th>
<th>RMSEA (probability ≤0.05)</th>
<th>Chi Square (degrees of freedom)</th>
<th>Significance Chi Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model</td>
<td>37.72/54.13</td>
<td>0.74</td>
<td>0.76</td>
<td>0.40</td>
<td>128.59 (23)</td>
<td>0.00</td>
</tr>
<tr>
<td>Quadratic model</td>
<td>28.27/46.05</td>
<td>0.77</td>
<td>0.78</td>
<td>0.39</td>
<td>117.14 (22)</td>
<td>0.00</td>
</tr>
<tr>
<td>Piece wise model</td>
<td>-25.31/-6.17</td>
<td>0.90</td>
<td>0.900</td>
<td>0.26 (0.00)</td>
<td>61.56 (21)</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Notes. N = 29; AIC – Akaike Information Criterion, BIC Bayesian Information Criterion: the lower the value, the better fit. CFI – Comparative Fit Index / TLI – Tucker Lewis Index: above 0.900 (preferably >0.98) is closest to the model fit. RMSEA <0.08 (preferably even <0.05) is ideal. Nonsignificant Chi Square Test: model fits best (no significant differences)

As can be seen in Table 8, the quadratic model has an almost equal fit (CFI = 0.77, TLI = 0.78, RMSEA = 0.39) as the linear model (CFI = 0.74, TLI = 0.76, RMSEA = 0.40). This is in line with the results of the RMA. In both model fits, the model fit is moderate. In line with our theory, an effect in the first 15 minutes (0min, 5min, 10min, 15min) was expected. With this theory in mind, a Piecewise Latent Growth Model is executed in Mplus, with a cut off at 15 minutes (slope 1 = 0, 5, 10, 15 minutes; slope 2 = 30, 45, 60 minutes), based on our hypothesis. This Piecewise Latent Growth Model has the best fit with the data (CFI = .90, TLI = .90, RMSEA = .26), although the RMSEA value (>0.08) indicates that there remains potential for improvement. The Piecewise model shows that the significant change takes place in the first 15 minutes, and afterwards stays stable.
Figure 8. Graphic representation of LGM linear model, LGM quadratic model and Piecewise LGM.

Comparing LGM and RMA

Comparing the SUD-scores of the RMA and de factor scores of the SUD-scores, see Table 9, there are three aspects worth mentioning. First, the overall trend that both a linear and a quadratic model could fit is the same in both statistical methods. This gives more robustness to the findings. Second, the variance of de factor scores is smaller than the variance of the RMA SUD-scores. This can be attributed to the filtered statistical noise by using factor scores. With a smaller variance, results are more precise. Third, the means of the factor scores are structurally lower than the means of the SUD-scores of the RMA. This can be explained by including six sensations in the LGM, instead of the seven sensations included in the RMA.
Table 9. Means and Variances of RMA and Factor Scores

<table>
<thead>
<tr>
<th></th>
<th>SUD scores</th>
<th>Factor Scores SUD scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeated Measures</td>
<td></td>
<td>Latent Growth Model</td>
</tr>
<tr>
<td>Analysis</td>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td></td>
<td>(variance)</td>
<td>(variance)</td>
</tr>
<tr>
<td>0min</td>
<td>1.11 (.44)</td>
<td>1.13 (.20)</td>
</tr>
<tr>
<td>5min</td>
<td>1.25 (.47)</td>
<td>1.17 (.18)</td>
</tr>
<tr>
<td>10min</td>
<td>1.32 (.49)</td>
<td>1.29 (.15)</td>
</tr>
<tr>
<td>15min</td>
<td>1.40 (.60)</td>
<td>1.30 (.27)</td>
</tr>
<tr>
<td>30min</td>
<td>1.44 (.78)</td>
<td>1.37 (.40)</td>
</tr>
<tr>
<td>45min</td>
<td>1.47 (.92)</td>
<td>1.36 (.48)</td>
</tr>
<tr>
<td>60min</td>
<td>1.39 (.92)</td>
<td>1.37 (.47)</td>
</tr>
</tbody>
</table>

Furthermore, in both RMA and factor scores the variance increases over time, this can been seen too in the graph of observed individual values. The unique trajectory of every case, with an intercept and slope, and individually development over time is depicted (Figure 9).

![Figure 9. Observed individual values.](image_url)
Repeated Measures Analysis Between Session Habituation

The data (mean SUD-scores of ERP 1 - ERP 10) are checked on normality and sphericity.

_Normally distributed data:_ to inspect the normality, the P-P plots (probability-probability plots), histograms of the standardised residuals and the values of skewness and kurtosis are checked. The data of the different measurement moments are normally distributed, so the assumption of normally distributed data is met.

_Sphericity:_ with Mauchly’s test in SPSS the hypothesis that the variances of the differences between conditions are equal is tested. Mauchly’s test indicated that the assumption of sphericity had been violated $\chi^2 (44) = 116.45, p < .01$, therefore the Greenhouse-Geisser estimates of sphericity are reported ($\varepsilon = .47$).

The results of the Repeated Measures Analysis describe the development of mean SUD-scores across sessions (see Appendix Table 10). It was found that the mean SUD-scores not significant differ across sessions, $F (4.26, 98.09) = .40, p > .05$. The second hypothesis, that the perceived severity of premonitory sensations will decrease across sessions (between-session habituation), is not confirmed.

![Graph](image)

*Figure 10. Graphic representation of the between-session habituation. Mean sensation severity scores (SUD-scores) across sessions.*
Figure 11. Close-up of the between-session habituation. Mean sensation severity scores (SUD-scores) across sessions.
Discussion

In this study, habituation as underlying mechanism in Exposure and Response Prevention (ERP) treatment in Tourette is tested. During ten one-hour treatment sessions, the premonitory urge severity is rated at 0, 5, 10, 15, 30 and 60 minutes. The main hypothesis was that there is a quadratic habituation trend to premonitory urges. An intensification of sensory sensations and urges was expected with tic suppression, followed by a decrease of severity of premonitory sensations, during the treatment sessions (within-session habituation). The results of both Repeated Measure Analysis and Latent Growth Model show that the SUD-scores significantly differ during sessions, with both a significant linear and quadratic effect. Besides LGM Linear and Quadratic curves, also a Piecewise Growth Model was tested, because of expected effects in the first 15 minutes. Of all models, this Piecewise Latent Growth Model has the best fit with the data. This Piecewise model shows that the significant change takes place in the first 15 minutes, and afterwards stays stable. This leads to the interpretation that during the first 15 minutes it is hard for the patients to suppress the tics and to endure the unpleasant sensation, and the severity of the experienced premonitory urges changes during this time period. After this difficult period of exposure and response prevention, there is a stabilisation of the unpleasant premonitory urge.

This finding does not support the hypothesis of this study, that within a session there is a significant reduction in severity of premonitory urge, following a quadratic habituation trend. Although a significant quadratic trend is found, there is no reduction in severity of premonitory urge within the sessions. But in line with our hypothesis, results show that during the first 15-minutes a significant increase takes place. Instead of a following reduction, like the linear reduction in the study of Verdellen et al. (2008), the initial increase is followed by a stable pattern. This stable pattern is also found in Specht et al. (2013). The current results suggest that tic suppression in ERP does result in a change in premonitory urge ratings within the sessions, but does not result in a reduction in this ratings or habituation. As reported by Van Minnen and Foa (2006), based on their findings of research to habituation in ERP treatment of posttraumatic stress disorder (PTSD), within-session habituation may not be a necessary condition for successful treatment.

The present results show that during the first 15 minutes of the ERP sessions (with a duration of 60 minutes) a process of change in premonitory sensations takes place, while the following 45 minutes are characterized by a stable pattern. The pattern of the SUD-score
within sessions leads to the inference that the first quarter of the sessions is of considerable importance. This is important information concerning the duration of sessions. Earlier research showed that ERP sessions could be shortened from two hours (Verdellen et al., 2004) to one-hour sessions (van de Griendt et al., in preparation). In line with this research, findings of research of Van Minnen and Foa (2006) in exposure treatment of PTSD demonstrate that 30-minutes exposures sessions are as effective as 60-minutes exposure sessions.

The current findings suggest that shortened ERP sessions, for example 15-minute sessions, may be as effective as the current sessions of 60 minutes, because the important change in SUD-score takes place in the first 15 minutes and afterwards the SUD-score stays stable. A clinical implication of this results is that shorter ERP sessions are easier to implement in clinical practice. In comparison to 60-minutes sessions, 15-minutes sessions are less expensive, less intensive and burdensome to patients as well as therapists, shorter sessions will possible have less dropout and will have a lower threshold for patients to start the treatment. This advantages contribute shorter ERP sessions to compete with the pharmacological treatment for Tourette. Further research will be required to prove if 15-minutes ERP sessions are as effective as 60-minutes ERP sessions in the treatment of tic disorders.

The second hypothesis was that there is a significant decrease in severity of premonitory sensations (SUD-score) across sessions (between-session-habituation). The results of the Repeated Measures Analysis indicate that there is not a significant difference in SUD-score across sessions. So in our study, there is no statistical support for reduction of sensation severity ratings between the ten ERP sessions. In accordance with this result, there was no significant difference between the mean score at the Premonitory Urge for Tics Scale (PUTS) before and after treatment. So analyses of SUD-scores as well as PUTS-scores indicate that there is no habituation across treatment sessions.

Despite of the lack of between-session habituation, a significant reduction of tic behaviour is found. Patients had a significant higher global tic severity score (measured with the YGTSS) before treatment than at the end of the treatment. This is in line with findings of previous studies, who also found that exposure and response prevention treatment can decrease tic behaviour (Hoogduin, Verdellen & Cath, 1997; Verdellen et al., 2004; Verdellen et al., 2008).
The finding of no between-session habituation deviates from the results of Verdellen et al. (2008), who did find a significant reduction of SUD-scores across sessions. Furthermore, the finding of no between-session habituation is not consistent with the theory that exposure and response prevention will cause habituation to the premonitory urges (Bliss, 1980; Hoogduin, Verdellen & Cath, 1997; Leckman, Walker & Cohen, 1993). This results run counter to the urge habituation hypothesis as an explanation for the ameliorative effects of tic suppression treatment ERP. Possibly the habituation to the unpleasant premonitory urge may not be the best explanation for how behavioural treatment results in symptom improvement; there might be other working mechanisms in the process of ERP treatment. A possible theory for this finding is that a cognitive process could take place. Instead of the habituation to the unpleasant premonitory urge, patients learn and experience during exposure and response prevention that they can control the tics and endure the premonitory sensations for prolonged periods. This experience can cause a falsification of their cognitions about the tic behaviour and the ability to endure and control the tics and urges. This cognitive process during ERP treatment can contribute to the reduction in tic behaviour. So possibly, there is a cognitive change (“I can control the tics; the feared tics will not happen”) instead of a physical habituation to the urges (to get used to the premonitory sensations).

This is in line with recent theories and rationale of anxiety disorders and obsessive compulsive disorder, where instead of the habituation-model the falsification-model is chosen as explanation mechanism for the effects of ERP treatment (Broeke, Korrelboom & Verbraak, 2009; Craske & Mystkowski, 2006; Verbraak, Hoogduin & Keijsers, 2011). The falsification-model states that during exposure treatment, patients learn that the feared consequences will not happen. The unreal expectations are tested and falsified, and new, alternative expectations are composed. During this process in ERP, the focus is at a change in expectations, instead of a decrease in anxiety or sensation. So, according to the falsification-model, the goal of ERP in tic disorders is to decrease the expectations that the patients can not control the tics and to create alternative expectations and cognitions that they can control the tics or that the feared tics will not happen. By preventing the tics from coming through, and exposing themselves to the premonitory urges that accompany the tics, patients learn that they can control their tics. This will result in tic reduction.

A theoretical implication of this finding is a possible modification of the current habituation theory of tic disorders, with the possible addition of a cognitive component as
underlying mechanism in exposure and response prevention treatment. To gain more insight into the nature of the underlying mechanisms of ERP in Tourette, further research is needed.

Several issues need to be addressed. As mentioned by Verdellen et al. (2008), this study only included a subjective measurement of the severity of sensations and urges. Addition of a more objective measurement, for example measurement of physical arousal, would possibly strengthen the reliability of the subjective reported SUD-score. However, an objective instrument for measuring urges is not known yet. Secondly, there are several properties of SPSS that make a less model fit with our data (Field, 2009; see ‘Statistic Analyses’). Besides, the study population consists of 29 participants. This is a relative small sample size, which makes it hard to attain statistical significant or meaningful results as well as sufficient coverage and power. However, from each participant there is information of 12 sessions and seven measurement moments within a session of eight urges, which makes the power larger and the findings more robust. The use of CFA’s and LGM in Mplus besides SPSS responds to the limitations of small sample size and the insufficient properties of SPSS concerning the data, and it gives more robustness to the findings.

To conclude, further research is needed to replicate these findings and extend them to other clinical populations. As mentioned earlier, more research is needed to draw conclusions on the optimal amount, frequency and length of the ERP sessions and to answer the question if shortened ERP sessions are as effective as current one-hour sessions. To gain more insight in the underlying working mechanism of ERP in tic disorders, more research is needed in the process of habituation and the role of cognitive aspects in this process.
Acknowledgment

I would express my appreciation and thanks to my supervisor Professor Dr. I.H.A. (Ingmar) Franken, Erasmus University Rotterdam, Faculty of Social Sciences, Department of Psychology, for his support, guidance and encouragement during my research internship. In favour of my fascination for the Syndrome of Gilles de la Tourette, he shared his contacts in the research field of tic disorders and introduced me to several Tourette experts. Thanks a lot for your advice and effort through the learning process of this master thesis as well as for the opportunities for my career.

I would like to express my warm thanks to my supervisor Drs. J.T.M. (Jolande) van de Griendt, HSK Den Bosch Expertise Tics, who supported me throughout the research internship. I am thankful for her inspiring guidance, constructive comments and suggestions, and the nice cooperation during this period. I respect her perseverance in the Summerschool Course ‘Introduction to Structural Equation Modelling using Mplus’ at the Utrecht University. Jolande, I enjoyed working together at this research project!

I also would like to acknowledge the statistical support of Dr. A.G.J. (Rens) van de Schoot (Utrecht University) in the Multilevel Statistics. His enthusiasm, humour and involvement facilitated the complex analyses and made statistics almost nice.

The research internship opportunity I had with HSK Den Bosch Expertise Tics was a great chance for learning and professional development. I am thankful for having this chance and I want to express my special thanks to HSK Den Bosch, especially to the head of Expertise Tics Centre Dr. C.W.J. (Cara) Verdellen, for the warm welcome and pleasant period at HSK.

Many thanks to my friends at Waalwijk for their hospitality and friendship during my research period at HSK Den Bosch. I appreciate your kindness, cosiness and warm interest. I had a very good time!

Finally, I would like to thank my beloved family and boyfriend, who have supported, loved and encouraged me throughout the entire process of this master thesis. I love you!
References


Appendix

Main Hypothesis

Table 2. Mean Scores Premonitory Urge for Tics Scale (PUTS)

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>27</td>
<td>23,30</td>
<td>5,66</td>
</tr>
<tr>
<td>M1</td>
<td>29</td>
<td>23,59</td>
<td>5,61</td>
</tr>
<tr>
<td>M2</td>
<td>27</td>
<td>24,13</td>
<td>5,01</td>
</tr>
<tr>
<td>M3</td>
<td>27</td>
<td>23,52</td>
<td>5,17</td>
</tr>
<tr>
<td>M4</td>
<td>13</td>
<td>22,69</td>
<td>6,32</td>
</tr>
</tbody>
</table>

Table 3. Statistics Paired Samples T-Test Pre- and Posttest score Yale Global Tics Severity Scale

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>YGTSS M0</td>
<td>26</td>
<td>18,81</td>
<td>7,28</td>
<td>1,43</td>
</tr>
<tr>
<td>YGTSS M3</td>
<td>26</td>
<td>12,54</td>
<td>6,80</td>
<td>1,33</td>
</tr>
</tbody>
</table>
Table 4. Paired Samples T-Test Pre- and Posttest score Yale Global Tics Severity Scale

<table>
<thead>
<tr>
<th>Paired Differences</th>
<th>95% Confidence Interval of the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>YGTSS M0 0-55</td>
<td>0.55</td>
</tr>
<tr>
<td>YGTSS M3 0-55</td>
<td>6.27</td>
</tr>
</tbody>
</table>

Table 5. Descriptives Statistics mean SUD-scores

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>meanSUD_0min_5sens</td>
<td>29</td>
<td>1.11</td>
<td>.66</td>
</tr>
<tr>
<td>meanSUD_5min_5sens</td>
<td>29</td>
<td>1.25</td>
<td>.68</td>
</tr>
<tr>
<td>meanSUD_10min_5sens</td>
<td>29</td>
<td>1.32</td>
<td>.70</td>
</tr>
<tr>
<td>meanSUD_15min_5sens</td>
<td>29</td>
<td>1.40</td>
<td>.78</td>
</tr>
<tr>
<td>meanSUD_30min_5sens</td>
<td>29</td>
<td>1.44</td>
<td>.88</td>
</tr>
<tr>
<td>meanSUD_45min_5sens</td>
<td>29</td>
<td>1.47</td>
<td>.96</td>
</tr>
<tr>
<td>meanSUD_60min_5sens</td>
<td>29</td>
<td>1.39</td>
<td>.96</td>
</tr>
</tbody>
</table>
Table 6. Test of Within-Subjects Effects within-session habituation

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power &lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor1 Sphericity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assumed</td>
<td>2.73</td>
<td>6</td>
<td>.45</td>
<td>5.81</td>
<td>.000</td>
<td>.17</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>2.73</td>
<td>1.49</td>
<td>1.83</td>
<td>5.81</td>
<td>.011</td>
<td>.17</td>
<td>2.73</td>
</tr>
<tr>
<td></td>
<td>2.73</td>
<td>1.55</td>
<td>1.76</td>
<td>5.81</td>
<td>.010</td>
<td>.17</td>
<td>2.73</td>
</tr>
<tr>
<td>Lower-bound</td>
<td>2.73</td>
<td>1.00</td>
<td>2.73</td>
<td>5.81</td>
<td>.023</td>
<td>.17</td>
<td>2.73</td>
</tr>
<tr>
<td>Error (factor 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sphericity Assumed</td>
<td>13.15</td>
<td>168</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greenhouse-Geisser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.15</td>
<td>41.63</td>
<td></td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Huynh-Feldt</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13.15</td>
<td>43.42</td>
<td></td>
<td>.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower-bound</td>
<td>13.15</td>
<td>28.00</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Test of Within-Subjects Contrasts within-session habituation

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
<th>Observed Power &lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor1 Linear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quadratic</td>
<td>2.03</td>
<td>1</td>
<td>2.03</td>
<td>5.50</td>
<td>.026</td>
<td>.16</td>
<td>2.03</td>
</tr>
<tr>
<td>Cubic</td>
<td>.65</td>
<td>1</td>
<td>.65</td>
<td>14.71</td>
<td>.001</td>
<td>.34</td>
<td>.65</td>
</tr>
<tr>
<td>Cubic</td>
<td>.02</td>
<td>1</td>
<td>.02</td>
<td>8.80</td>
<td>.378</td>
<td>.03</td>
<td>.020</td>
</tr>
<tr>
<td>Cubic</td>
<td>2.03</td>
<td>1</td>
<td>2.03</td>
<td>5.50</td>
<td>.026</td>
<td>.16</td>
<td>2.03</td>
</tr>
<tr>
<td>Error (factor 1) Quadratic</td>
<td>10.34</td>
<td>28</td>
<td>.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.23</td>
<td>28</td>
<td></td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubic</td>
<td>.68</td>
<td>28</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 7. Linear, quadratic and piecewise LGM model of the within-session habituation
Second Hypothesis

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>factor1</td>
<td>Sphericity Assumed</td>
<td>,95</td>
<td>9</td>
<td>,11</td>
<td>,40</td>
<td>,936</td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>,95</td>
<td>4,26</td>
<td>,22</td>
<td>,40</td>
<td>,822</td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>,95</td>
<td>5,36</td>
<td>,18</td>
<td>,40</td>
<td>,862</td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>,95</td>
<td>1,00</td>
<td>,95</td>
<td>,40</td>
<td>,535</td>
</tr>
<tr>
<td>Error (factor 1)</td>
<td>Sphericity Assumed</td>
<td>55,16</td>
<td>207</td>
<td>,27</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Greenhouse-Geisser</td>
<td>55,16</td>
<td>98,09</td>
<td>,56</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Huynh-Feldt</td>
<td>55,16</td>
<td>123,20</td>
<td>,45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lower-bound</td>
<td>55,16</td>
<td>23,00</td>
<td>2,40</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>