“Did not observe”
Reflecting On and Improving the Measurement of Young Children’s Social and Emotional Competence
“DID NOT OBSERVE”
REFLECTING ON AND IMPROVING THE MEASURES OF YOUNG CHILDREN’S SOCIAL AND EMOTIONAL COMPETENCE

Fanny Gulliksson

This thesis concerns ways to measure social and emotional competence with high quality. The utility of the Social Competence Scale – Observer (SCO) was evaluated and compared with Social Competence Scale – Teacher (SCT). Both consisting of the sub-scales Prosocial/communication Skills Scale and Emotion Regulation skills Scale. Data arrived from the child intervention program PsPATHS. One sample of 109 preschoolers was rated on SCO during play observations of which 101 children were also rated by teachers using SCT. The utilization of SCO showed both consistency and inconsistency compared with previous measures regarding reliability, depending on subscale. The comparison between SCT and SCO correlated significantly on the Prosocial/communication Skills Scale. SCT had excellent internal consistency reliability on the same subscale. Observers frequently reported “not observed” behaviors suggesting that items should be reworded or replaced. Additional measurement methods and repeated evaluations of psychometric properties are recommended for future measures of children’s’ social and emotional competence.

It is widely agreed that the development of social and emotional competence is fundamental in order to establish and maintain meaningful relationships (Bierman et al., 2008; Denham, 2006; Rose-Krasnor, 1997). These competencies are regarded as essential for physical and psychological well-being among young children starting from infancy (Bierman et al., 2008; Denham, 2006; Rose-Krasnor, 1997). Conversely, insufficient social and emotional competence has been related to poor social outcomes, such as peer rejection and adjustment difficulties in school settings (Greenberg, 2006) and later risk behaviors like anti-social behaviors and academic difficulties (Denham, 2006; Silk, Steinberg, & Sheffield Morris, 2003).

Given these circumstances, many resources have invested on large scale child intervention programs (Denham, 2006). Within this type of intervention, there is an increasing need to gain insight into appropriate measurement methods capturing social and emotional competencies among young children (Zaslow et al., 2006). Among scholars, there is also a necessity for a unified understanding of social and emotional competence in order to develop these measurements (Cordier et al., 2016; Rose-Krasnor, 1997). Furthermore, most scales and instruments are sensitive to context, and reviews of their psychometric properties have to be replicated several times in order to establish validity and reliability (Cordier et al., 2016). In the present study, two versions of Social Competence Scale; the Observer Version (SCO) and the Teacher version (SCT), developed by the Conduct Problem Prevention Research Group [CPPRG] (1995) were compared and examined. SCO and SCT assess prosocial/communication skills and
emotional regulation, which are regarded as major components of social and emotional competence.

**PsPATHS and PATHS**
The two versions of Social Competence Scale; SCO and SCT, were used in the Swedish adaptation and intervention trial of PsPATHS (Promoting Preschool Alternative Thinking Strategies Pre-school edition). The intervention program was originally developed in the United States with the aim to increase school readiness by strengthening social and emotional competence among pre-school children, regardless of socio-economic background (Bierman, Nix, Greenberg, Blair, & Domitrovich, 2008; Conduct Problems Prevention Research Group, 2010; Domitrovich, Cortes, & Greenberg, 2007). Apart from early literacy, school readiness is referred to the child’s capacity to form positive peer and teacher relationships which in turn is derived from the child’s social and emotional competence (Bierman et al, 2008).

PsPATHS was incorporated into *Head Start*, the largest early childhood education program in U.S. that especially targets children from disadvantaged socio-economic backgrounds (Kelly, Edgerton, Graham, Robertson, & Syme, 2015). This way it could be tested through two randomly controlled trials with children (245 and 356 children in each trial) aged three to five who were enrolled in Head Start Centers. Both studies found evidence of increased social and emotional competence among participating children (Bierman et al, 2008; Domitrovich et al., 2007).

PsPATHS was developed for children enrolled in pre-schools, whereas the original curriculum, PATHS (Promoting Alternative Thinking Strategies) was geared toward older children who were enrolled in elementary schools. In a two year cluster randomized trial of the elementary school-based version, PATHS, it was confirmed that social and emotional competence increased among a diverse sample of children, after participating in the program. Furthermore, authors suggested that aggressive behavior was significantly reduced compared with control groups (Crean, & Johnson, 2013). Consequently, both PATHS and PsPATHS have been, and are currently being, adapted and implemented in diverse English and non-English speaking countries such as the Netherlands, Turkey, Croatia, England, Canada, Thailand, Hong Kong and Sweden (Inam, Tariq, & Zaman, 2015).

**Social and Emotional Competence**
Social and emotional competence consists of a variation in distinct, yet related, skills which can appear within numerous contexts, with multiple individuals in different interactions (Cordier et al., 2015). A feasible definition regarding emotional competence is that it involves the ability to understand one’s own and others’ feelings in the surrounding context and to be able to manage them (Halle, & Darling-Churchill, 2016). The emotion management consists of internal and external processes that evaluates and monitors emotional reactions and is commonly defined as Emotional regulation (Kopp, & Neufeld, 2003). Emotional regulation more specifically involves being able to increase or decrease physiological arousal and has been regarded important in terms of how children might resist peer pressure and control their impulses or be able to reflect on their feelings (Kostelnik, Stein, Whiren, Soderman, & Gregory, 2002). Emotional regulation been significant correlated with long term effects of social competence (Denham et al., 2003) showing how these constructs interacts with each other.
The complexity of social competence has in part been explained by Rose-Krasnor (1997) who perceived that the term “social competence” had been used too implicitly in literature as a construct everyone agreed on. However, measures of social competence have on contrary shown large variations, not only between measurement methods, but also between assessors using the same instrument (Cordier et al., 2016; Denham, 2006).

One approach to describe social competence has been by pro-social behaviors defined as voluntary behaviors aiming to benefit others (Eisenberg, & Fabes, 1998). Although this seems like an easy and comprehensive definition, pro-social includes a multifaceted spectrum of different perspectives such as socialization perspectives and perspectives of context and culture. It has been shown that desired behaviors, such as prosocial skills, might shift depending on cultural norms (Han, & Kemple, 2006). Global behaviors of pro-social skills have however been related to traits of helping behaviors, responsiveness and the ability to work with others (Bierman et al., 2008). Children’s tendency to show such behaviors has been related to how well they knew the person they interacted with and how motivated they were in doing it (Eisenberg, 1983). Pro social skills are regarded to impact the degree of effective social interactions and behaviors related to establishing relationships. Prosocial skills have therefore together with emotional regulation been considered important to increase through interventions such as PsPATHS (Bierman et al., 2008).

Social and emotional competence develop in pace with the child’s age and maturity level (Bierman et al, 2008; Kopp & Neufeld, 2003). By the age of four, children have usually established relations outside of the family, such as friendships with peers and pre-school teachers. At this age, the comprehension of socially appropriate and unacceptable behaviors generally increases, such as knowing that sharing is nice. Most children, by this age, manage to control emotions rather well and can come up with strategies for problem solving (National Association for the Education of Young Children, 2009). It is during this period of development that intervention programs such as PsPATHS aim to support social and emotional competence promotion. They do so through a structured curriculum that has a broad theoretical base and involves practices used by established pre-school teachers. The curriculum aim to specifically target constructs such as prosocial skills (Domitrovich et al., 2007). It is crucial to remember however that the majority of research on social and emotional competence has been conducted in studies of children in Western societies, especially North American studies. A growing research body suggests that social and emotional competencies are not equally important or expressed the same in all cultures and scholars have stressed the need to explore these constructs in a broader field of culture (Han, & Kemple, 2006; Halle, & Darling-Churchill, 2016).

Measuring Social and Emotional Competence

As a consequence of the broad and complex definition of social and emotional competence, measurements of these constructs are regarded equally complex. In order to track the success of intervention programs, scholars has empaties the need to carefully consider and evaluate why, how and when assessment tools are most appropriately used (Cordier et al., 2016; Denham, 2006).

In contexts of large scale interventions, the usability of the instrument and its administration are regarded with importance, which means that practitioners of assessment need to find the instrument useful and understandable (Halle, & Darling-Churchill, 2016). Behavior-rating scales have commonly been used for assessments of
social and emotional competence and have generally been conducted by teachers and parents who rate the competencies of children based on their experiences (Matson & Wilkins, 2009). The method is considered inexpensive and user friendly. However, assessment instructions needs to be clarified for practitioners with regard to the difficulty to rate abstract attributes such as social and emotional (Kline, 2009). Besides, rating scales has been connected with rater bias and the risk of making subjective assessments (Cordier et al., 2016; Spinrad, Eisenberg, & Gaertner, 2007).

Behavior-rating scales are based on observed behaviors. Methods of observations have been discussed differently depending on abstraction level. On a molecular level, observers decoded fine-grained details like “opening mouth” and “squeezing eyes”. On a more global and molar level, observers instead created distance to capture the meaning of combined molecular units into a more overall category such as “yawning” or ever larger “being tired” (Miller, 2007).

In child studies, observation methods also vary on a spectrum from being very systematic and standardized to more flexible and naturalistic. The selection of an appropriate method is usually depending on the research question, behaviors of interest and the occurrence rate of these (Miller, 2007). Systematic methods with a fixed time coding and schedule are usually regarded with a high objectivity and more associated with validity and reliability (Shapiro, & Skinner, 1990). However, the use of standardized observation methods for behaviors with low occurrence rate has also been correlated with insufficient amount of data (Miller, 2007). Additionally, longitudinal intervention studies, involving several measurements, are seen with high risks of dropout’s rates (Howell, 2013). So when observational methods are planned in longitudinal child interventions, feasibility to access rich data is regarded crucial.

A recommended option in cases of subtly and less frequent behaviors, have instead been to use more flexible methods, facilitated by individuals who are part of in the naturalistic setting. Concerning children’s behaviors, these individuals are usually parents and teachers performing narrative recoding, taking notes of behaviors as they happen (Miller, 2007). But as mentioned, although the rate of behaviors may increase, teacher ratings has been regarded as more sensitive to bias and contradictory with the additional risk of missing behaviors when taking notes (Skinner, Freeland, & Shapiro, 2003).

In the field of controlled observations, recommended guidelines have developed. Harris and Lahey, (1982) empathized to keep observers independent during data collection with restricted communication between each other in order to reduce coding bias. Equally important was to define the construct of interest in specific behaviors, and that coders were provided with a thorough introduction (Smith, 1986). Frequently conducted reliability checks were recommended as well (Harris, & Lahey, 1982; Smith, 1986). However, the use of very systematic methods has been under debate. Farmer-Dougan and Kaszuba (1999) highlighted the benefits by observing children in their natural setting, in which teachers could create and stimulate play in order to screen for cognitive and social competencies. Comparing it with more traditional standardized methods their findings suggested that the method was equally valid. Researchers are also more recently paying attention to the importance of ecological validity, and play assessments (Denham, 2006)
Regardless of measurement preference, researchers have been in agreement, stating that one single instrument is not sufficient for measuring all aspects of social and emotional competence in young children (Cordier et al., 2016; Denham, 2006, Halle, & Darling-Churchill, 2016, Rose-Krasnor, 1997). Instead, it has been recommended to use a mix of several assessment methods, such as merging results of behavior rating scales with measurements of physiological changes and coded observations (Cole, Martin & Dennis, 2004).

Psychometric properties and handling missing data
Looking at the complexity of measuring and defining social and emotional competence, scholars have emphasized the importance of psychometric properties (Cordier et al., 2016). Regardless of research design, unreliable and invalid scales might hamper whole studies, causing misleading conclusions (Howell, 2012). In order for measurements to be reliable, scores on instruments need to show consistency, and the level of consistency can guide researchers to know if changes are needed in the instrument or measurement method (DeVellis, 2003). Kline (2009) addressed the issue of reliability induction and referred to the method of only reporting old reliability information thereby ignoring the value of repeated measurements.

There are different kinds of ways to measure validity and reliability. Carmines & Zeller, (1979) presented concurrent validity which is to confirm the validity of new tests by getting a high correlation with an already validated test. Concurrent validity has been regarded with a rather limited usefulness since abstract constructs within psychology are difficult to apply criterions to (Carmines, & Zeller, 1979; Kline, 2000). In observations with multiple observers, inter rater reliability have been useful to establish how similar two independent observers’ rate behaviors addressed as social and emotional competence (Kline, 2009). Another method has been to measure the internal consistency reliability, which normally is calculated with Cronbach’s alpha coefficient. Internal consistency has been explained as how well items in a measurement instrument fit together conceptually and how they correlate with each other. The accepted value range from .7 up to 1 depending on what the measure will be used for (Kline, 2009).

Studies of research involving quantitative data are almost never totally complete which depending on amount and structure missing data might be very problematic (Howell, 2012). Missing data can derive for several reasons and in different amount. It might include missing data on one item or one observation or multivariate, missing data on several units (El-Khoury, 2012). Missing at random includes data where participants have dropped out randomly or just been away during assessments. Random missing data is regarded less problematic as long as there is a normal evenly distribution of missing cases, but is seen to reduce power (Howell, 2012). Systematic missing data is however more problematic and can cause severe problems when conducting analysis since it might include bias (El-Khoury, 2012; Howell, 2012). For this reason it has been important to evaluate what kinds of missing data a study includes and the frequency of it on order to later decide how to handle it. In some cases, replacing missing values by imputation has been valid than exclusion (Howell, 2012).

Social Competence Scale
The SCS includes items such as: “expresses needs and feelings”, “very good at understanding feelings”, “controls temper”, “acts friendly towards others” and “resolves peer problems on own” The scale items are very similar to Denham’s theoretical view of what social and emotional competencies involves: emotional expressiveness,
understanding of emotions, regulation of emotions, social and relationship skills, and social problem solving (Denham, 2006).

The SCS has been used in earlier implementations of the PsPATHS in the U.S. (Bierman et al., 2008). The scale was originally developed by a project called Fast Track, a longer and more complex intervention program (included PATHS but also had other interventions) with the aim of increasing social and emotional competence among beginning elementary school children (Corrigan, 2003; Conduct Problem Prevention Research Group (CPPRG) 1995). The original version of the SCS included 25 items which in turn was a consolidation of the existing scales “Health Resources Inventory” And Teacher Rating of Social Skills, as well as some newly developed items (Gesten, 1976; Kendall, & Wilcox, 1979 as cited in Corrigan, 2003). The scale has been tailored in four versions, one for parents (SCP), one for siblings, one for teachers (SCT) and one for independent observers (SCO). SCT and SCO are the versions of this instrument used in the Swedish trial of PsPATHS.

Within Fast Track, SCT was used by teachers with children aged four to seven from various ethnic backgrounds throughout the United States (CPPRG, 1995). Within PsPATHS in the U.S. the observer version, SCO was developed and tested for use during play observations (Bierman et al., 2008). This is also how it was used in the Swedish adaptation. Together with SCT and other measurements the scale served as abasis to evaluate the effects of the intervention (Bierman et al., 2008). When it was originally developed within Fast Track as a rating scale (SCT), an explorative factor analysis with Varimax rotation of the Social Competence Scale yielded a two factor-structure. The first factor was a combined factor that included Prosocial/communication Skills Scale and the Emotion Regulation Skills Scale, the other with was related to items regarding academic skills or readiness. The academic skills scale is not used in the play observations.

Emotional Regulation Skills Scale within SCO was evaluated during the pilot year of the Swedish adaptation of PsPATHS. In line with Corrigan (2013), this scale resulted in a one factor structure indicating that all items can be derived to one single construct (Gulliksson, 2014). Another study, within the same project looked at the factor structure of Emotion Regulation Skills Scale and Prosocial/communication Skills Scale which resulted also resulted in one facture structure (Kvamme, 2014). Missing data was reported to be critically high on some items. Especially item 1 “resolves peer problems on own”, Item 6 “Listens to others point of view” and item 9 “Copes well with failure” (Kvamme, 2014). In order to systematically track future frequencies of missing data, the additional option “did not observe”, was added on the rating sheet for behaviors observers regarded as nonexistent.

For the teacher version of the scale (SCT) the Cronbach α coefficient in the U.S. Fast Track trial was .98 for the total scale. The separate Cronbach α for each sub-scale: Prosocial/communication Skills Scale and Emotion Regulation skills Scale, separately was .96. This indicate that the instrument has high internal consistency, in the teacher rating edition, and is sometimes used both together as a total score combining prosocial skills and emotion regulation and individually as separate sub scales (Corrigan, 2003).

The intraclass correlation coefficient (ICC), explained as an intrarrater reliability measure, in which raters’ scores are compared, was .7 in Bierman et al.’s, study (2008) for the SCO or observational measure. This, according to guidelines, is perceived as
“good”. In two past studies, conducted with data from the Swedish pilot studies, the ICC between raters using SCO resulted in “excellent” values of .76 for the Emotion Regulation Skills Scale (Gulliksson, 2014), and .75 for the Prosocial/communication Skills Scale (Kvamme, 2014). (< .40 = “poor”, between .40-59 = “fair”, .60-.74 = ”good” and .75 - 1.00 = “excellent”, Cicchetti, 1994. In Bierman et al’s study (2008) SCT and SCO showed a positive statistical significant correlation (r = 0.29 p < 0.05). Previous correlation between play observation sessions during the pilot trial of the Swedish version of PsPATHS resulted in r = .53, p < 0.01.

Gouley, Brotman and Huang (2008) evaluated the Social Competence Scale using the parent version (SCP) SCS, two different samples of pre-school aged children, one with children enrolled in a preventive intervention trial and the other with children enrolled in universal public pre-schools. SCP was evaluated during two years measuring psychometric properties like internal consistency and facture structure, both found to be as adequate as in Bierman et al’s study (2008) and was regarded as relatively stable over the two years with Cronbach α ranging from .87 to .92 at different assessment points. With their findings, they supported future use of the SCS as a brief measure of social and emotional competence or as a screening tool in order to detect children with deficits in social skills. Finally, they recommended that it would be of interest to compare versions of SCS with each other (Gouley, et al, 2008). In this study, baseline data derived from ratings using SCT and SCO within the Swedish trial of PsPATHS was compared, as recommended by Gouley et al. (2008).

Social and emotional competence in large scale interventions has mainly been related to different kinds of skills and strategies such as emotional regulation and pro-social skills (Bierman et al., Gouley et al, 2008). Emotional competence has in previous studies with pre-school aged children been significantly correlated with long –term social competence and the two constructs are therefore regarded to be overlapping and interrelated in complex ways (Denham et al., 2003). In order to measure these competencies in adequate manners, scholars need to gain insight on appropriate measurement methods (Zaslow et al., 2006). One way to do so is by evaluating existing and used measurements.

The present study
The aim of this current study was to evaluate the sub-scales Prosocial/Communication Skills Scale and Emotional Regulation Skills Scale included in SCO and SCT with a special focus on SCO. The Emotional Regulation Skills Scale included in SCO has previously been evaluated by the author (Gulliksson, 2014).

For this present study, data derived from SCO on both sub-scales, consisted of new observers, and new sample. Research questions of interest are: 1. Exploring the impact new raters and a new sample have on the psychometric properties of SCO. Chosen analysis for this is to examine inter rater reliability between raters using intraclass correlation coefficient (ICC) and percentage of agreement between raters. A correlation between morning play observations and afternoon play observations will also be conducted to investigate if the timing of assessment has an impact on the observations. (Cronbach’s alpha will be measured to check internal consistency of SCT instead of ICC). 2. To find out how SCO are related to SCT, first by checking descriptive data and inter-item correlations. Second, by execute another correlation between ratings conducted by teachers, using SCT, and observers using SCO on the same sample for
scales that had shown good reliability. 3. How observers have used the option of “did not observe” which will be explored through frequency measures.

Method

Participants
The selection procedure within the Swedish adaptation of PsPATHS consisted of a random selection from three specific municipalities located in Stockholm County. Schools were invited to participate by e-mail or phone calls to check interest of participation. Pre-schools showing interest received visits from the research group who provided more information. For this particular study, data from the first pretest wave of PsPATHS was used; the PsPATHS trial in Sweden has two intervention waves, each with a pretest and posttest. The dataset covered the total selection of participating control and intervention schools enrolled at baseline during 2014. No school had then yet started the implementation of PsPATHS. Within each enrolled pre-school, children whose parents had given written consent participated in the play observation. In total, 129 children born 2009 and 2010 from 15 different pre-schools were enrolled in the first pretest. All children were supposed to participate in two play observations being rated by observers using SCO and rated once by their teacher using SCT.

Material
Two different play kits were used as material for the observations. Playmobil Country, Pony farm (5222) was used in morning observations. This miniature farm could be assembled and contained besides the farm plastic figures in shapes of people, animals, plants, tools and pieces of a fence that could be put together. The toy used in the afternoon was a ball track called Deluxe Marble race/ Marble Run Play Set, (WK-883-02). This plastic ball track consisted of small plastic pieces that could be built into different paths for marbles to run through.

Measures
The original version of Social competence Scale – Observer version (SCO) consisted of 25 items divided on three sub-scales; Prosocial/communication skills scale, Emotion Regulation skills scale and Academic Achievement Skills Scale (Corrigan, 2003). In the Swedish trial of PsPATHS, all eight items included in sub-scale used for prosocial behaviors was kept and used during pre-test wave one. From the Emotion Regulation skills scale, six out of ten items were used after being evaluated during the Pilot Year. The exclusion of items on the emotional regulation scale was partly due to cultural adaptation, because items such as “Can wait in line patiently” is a rare thing to learn in Swedish pre-schools, as lines are not common. No items from the Academic Achievement Skills Scale were used in SCO, nor were it used in Bierman et al’s study (2008). The 14 remaining items on Prosocial/communication skills scale and Emotion Regulation skills scale were rated on a 5-point Likert scale from 1 (“not at all,”) to 5 (“very well”) with the addition of “did not observe”. The items stated behaviors that children may express during play in pre-schools such as “Share materials”, “is helpful to others” and “copes well with failure”. SCO was kept in English for observers. Because all observers had adequate language skills there was no reason to translate it. (See Appendix C)

Social Competence Scale –Teacher version (SCT) consists of the same items as SCO. However, all three sub scales were used. In the Prosocial/communication skills scale,
item one “Resolves Peer problem on own” and item eight “Accept things not going well” was excluded in SCT after evaluations from the Pilot Year, based on psychometric analysis. SCT was translated into Swedish for Pre-schools teachers. Just like SCO, the Likert points that teachers could rate ranged from 0 (“Not at all”) to 4 (“very well”), although translated into Swedish (“Stämmer inte alls”, “Stämmer lite grann”, “Stämmer ganska väl”, “Stämmer mycket väl”). There is no did not observe option on the teacher survey for these scales.

Procedure for Observers
A total of eight observer (one male) were part of the assessment team during the baseline measurement used for this study. Among them, the male only participated as a rater four times. Without him, the mean number of observations for each rater was around 30. All observers were enrolled psychology students and one rater had been enrolled during the second pilot wave and had previous experiences. I, as author of this thesis, was not involved as an observer for this dataset but had participated as an assessor and observer during other waves of measurements conducting a total of around 300 observations and 150 assessments using other measurements included in the PsPATHS intervention trial. The decision to choose data from the first pretest was to increase objectivity and to have a selection of participating children at baseline before the intervention of PsPATHS started.

All new raters were given training before assessment by watching a video recorded play observation with two children. Afterwards, everyone rated each child based on a codebook conducted by the research group. The codebook was conducted after the first pilot trial. All items were discussed with supervisors of the research group. (Appendix D for codebook)

Each child participated in two play observations during the same day; once in the morning when the observers arrived and once during the afternoon. Between observations each child met with the observers individually conducting other assessments provided by the PsPATHS research team. Upon morning arrival, the observers divided participating children in each pre-school into groups of two or three. This was done with some collaboration with pre-school teachers in order to match attendance of the child as well as the curriculum of the pre-school. Sometimes sub groups of children within the pre-school were away on their own projects. The groups were formed with the consideration that all children in each group should know each other to some degree. Each child was informed that the play was voluntary and that they could leave the play if they wanted to. All were informed that the observers were visiting the pre-schools to learn more about how children play and think. The group of children that was going to be observed that particular day was asked to join in a play together.

Children were asked to explore a new toy brought by the observers. The play observations were conducted in a separate room provided by the pre-school so that it would not interfere or be disturbed by the pre-school’s daily activities. The toy was already set up when the group of children were introduced to the play time. They were asked to play together with the toy for 15 minutes and that the observers were going to tell them when the time was up. Observers did not correct children who started to play with something else; however it was noted on the rating sheets. In some occasions the observers interacted with the children, for example instructing them to “play nicely”
with the toys if they were playing rough, for example throwing things around, or that they should be careful with each other. In a few observations, pre-school teachers entered the room to get something, or requested to be present during the observation, however without interacting or talking.

During all observations two raters were present. They took notes during the observation, looking at all children at the same time. They were free to observe all children during the observation time instead of having a specific amount of minutes for each child. This was based on the idea that a more free observation would capture more interactions as well as each child’s behaviors. The play ended after fifteen minutes in which the observers interacted with the children again. In some cases, the observation had to be ended earlier, for example if a child needed to visit the bathroom or if they all had started to play with something else. After the observation, the observers rated each child individually using SCO with support from notes and codebook. Observers kept their ratings to themselves, never showing or discussing their ratings in order to ensure independency of raters.

**Teachers using SCT**
Pre-school teachers in each participating pre-school of PsPATHS rated participating children using SCT. The scale items were included in a rating booklet for each child provided by the assessment team of PsPATHS. Teachers were informed to send all booklets back to the Institution of Psychology using stamped envelopes given together with the booklet. Missing data excluded, a total of 101 children were rated with SCT during pre-test wave one. For this thesis, data from items included in the Prosocial/communication Skills Scale were used to compare teacher ratings with observer ratings using SCO. The child sample was the same for both. SCT for teachers was translated into Swedish.

**Ethical considerations**
The application to implement the Swedish adaption of PsPATHS trial has been approved by The Central Ethical Review Board and all ethical guidelines has been followed. Children were free to participate but in the concept of pre-schools, children are used to follow instructions. It was therefore important that observers were sensitive and responsive in cases where children seemed displeased.

**Analysis**
All entered data and analysis was performed using SPSS software. ICC was used to check inter rater reliability. ICC is a type of ANOVA that is commonly used to measure reliability between multiple raters. A two-way random model was chosen since observers were regarded as a random variable and could be exchanged by other observers. The difference between consistency was chosen over absolute agreement since close agreement between raters would have been excluded using absolute agreement. In interpreting the value, 0 is regarded as no agreement at all while 1 represents perfect agreement. (< .40 = “poor”, between .40-.59 = “fair”, .60-.74=“good” and .75-1.00= “excellent”, Cicchetti, 1994)

In regard to the high amount of missing data reported from the Swedish pilot study PsPATHS using SCO it was of high importance to investigate if missing data continued to frequently occur and in which patterns. With percentage of agreement patterns of how raters used “did not observe” could be visualized. This was calculated by coding
patterns between each rater for each child at each item and at each observational session. For example, when Items scored the same it was coded as 1, or when they differed on one score that was coded 2. All of these codes were created in a new variable for all 14 items multiplied with the two sessions of observation. With the experience of conducting these play observations it was also of interest to more closely check which items observers had difficulties observing and patterns of how they disagreed.

Result

Dropouts and missing data
A total of 129 children were enrolled in this sample of which eleven never participated in the play observation (SCO). These children were absent during the days that the observers were visiting the pre-school. There was another data loss between observation sessions. Although 118 children participated in the morning observation, nine of them had already gone home for the day before the afternoon observation was conducted resulting in a total of 109 children. Regarding teacher ratings (SCT) 101 of the remaining 109 children were rated. The teacher rating was voluntary and therefore some teachers decided not to conduct them. Drop outs excluded, missing data frequently occurs as “did not observe” and will be presented in own figures in the result section.

Reliability
The inter rater reliability between raters across observations sessions (morning versus afternoon) was conducted with 109 children using the Prosocial/communication skills Scale (SCO) and results showed good agreement [ICC] = .63 with 95 % confidence interval or CIs from .50-.74. On the Emotional regulation Skills Scale (N=109) the inter rater reliability between raters across observation sessions was instead found to be poor [ICC] = .24, 95% CIs [-.03-.46] based on ICC criterion for values reported by Cicchetti (1994). Due to this poor ICC value, Emotional regulation Skills Scale was not used in any additional analysis with the teacher rating scale, comparing the SCO and SCT. The correlation between SCT and SCO are therefore only represented with data from the Prosocial/communication Skills Scale. This also applies for the correlation between observations sessions.

Ratings for 101 children were obtained from teachers using Social Competence Scale – Teacher. The internal consistency was reported high, Cronbach α = .94. The range of Cronbach α if specific items were deleted was between .92 - .95.

In Table 1 and 2 it is seen that all items on SCT are normally distributed. The skewness and kurtosis lies within values of -1-.1. The standard errors for the skewness and kurtosis were, when multiplied by three, larger than skewness and kurtosis further showing that data were normally distributed. Regarding SCO, all items were normally distributed. Both SCO and SCT show close mean values for each item.
Table 1. Descriptive data for SCT showing arithmetic mean and standard deviation for all items. Prosocial/commination Skills Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Very good at understanding feelings</td>
<td>3.49</td>
<td>1.23</td>
<td>-.377</td>
<td>-.952</td>
</tr>
<tr>
<td>3. Shares material</td>
<td>.380</td>
<td>.92</td>
<td>-.388</td>
<td>-.613</td>
</tr>
<tr>
<td>4. Cooperate with friends without prompting</td>
<td>3.81</td>
<td>1.05</td>
<td>-.469</td>
<td>-.733</td>
</tr>
<tr>
<td>5. Is helpful to others</td>
<td>3.87</td>
<td>1.02</td>
<td>-.611</td>
<td>-.443</td>
</tr>
<tr>
<td>6. Listens to others point of view</td>
<td>.354</td>
<td>1.12</td>
<td>-.364</td>
<td>-.707</td>
</tr>
<tr>
<td>7. Acts friendly toward others</td>
<td>4.00</td>
<td>.97</td>
<td>-.672</td>
<td>-.520</td>
</tr>
</tbody>
</table>

Table 2. Descriptive data for SCO showing arithmetic mean and standard deviation for all items on Prosocial/commination Skills Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>M</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Resolves problems on own</td>
<td>3.63</td>
<td>.78</td>
<td>-.10</td>
<td>-.59</td>
</tr>
<tr>
<td>2. Very good at understanding feelings</td>
<td>3.51</td>
<td>1.02</td>
<td>-.33</td>
<td>-.61</td>
</tr>
<tr>
<td>3. Shares material</td>
<td>3.98</td>
<td>.79</td>
<td>-.87</td>
<td>.30</td>
</tr>
<tr>
<td>4. Cooperate with friends without prompting</td>
<td>3.59</td>
<td>.95</td>
<td>-.38</td>
<td>-.24</td>
</tr>
<tr>
<td>5. Is helpful to others</td>
<td>3.58</td>
<td>1.06</td>
<td>-.45</td>
<td>-.64</td>
</tr>
<tr>
<td>6. Listens to others point of view</td>
<td>3.68</td>
<td>.91</td>
<td>-.39</td>
<td>-.66</td>
</tr>
<tr>
<td>7. Acts friendly toward others</td>
<td>3.80</td>
<td>.75</td>
<td>-.35</td>
<td>-.66</td>
</tr>
</tbody>
</table>

SCT versus SCO
Pearson’s coefficient of correlation was calculated to compare the Social Competence Scale; teacher version (SCT) with the observer version (SCO) The arithmetic mean was calculated on scale level for each scale. For SCO (M = 3.63, SD = .79) and for SCT (M = 3.75, SD = .93) with a total N = 90 and showed statistical significance (r = .27, p < .05 (2-tailed))

Table 3 and 4 shows that all correlations between items on SCT were statistically significant at the .01 level (two tailed). (Note that item 1 “Resolves peer problems on own” was not included in SCT). For SCO each item had a mean value that was merged across observers and observation sessions. All items except showed a positive statistical significance at the.01 level (two tailed).
Table 3. Correlation matrix between items on Social Competence Scale – Teacher, using Pearson’s correlation (two tailed).

<table>
<thead>
<tr>
<th>Item</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very good at understanding feelings</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shares material</td>
<td>3</td>
<td>.68**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperate with friends without prompting</td>
<td>4</td>
<td>.62**</td>
<td>.81**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is helpful to others</td>
<td>5</td>
<td>.67**</td>
<td>.77**</td>
<td>.82**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listens to others point of view</td>
<td>6</td>
<td>.72**</td>
<td>.79**</td>
<td>.79**</td>
<td>.85**</td>
<td></td>
</tr>
<tr>
<td>Acts friendly toward others</td>
<td>7</td>
<td>.58**</td>
<td>.71**</td>
<td>.77**</td>
<td>.74**</td>
<td>.76**</td>
</tr>
</tbody>
</table>

** p < .01

Table 4. Correlation matrix between items on Social Competence Scale – Observer using Pearson’s correlation (two tailed).

<table>
<thead>
<tr>
<th>Item</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolves peer problems on own</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very good at understanding feelings</td>
<td>2</td>
<td>.55**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shares materials</td>
<td>3</td>
<td>.41**</td>
<td>.67**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cooperate with friends without prompting</td>
<td>4</td>
<td>.57**</td>
<td>.59**</td>
<td>.50**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is helpful to others</td>
<td>5</td>
<td>.38**</td>
<td>.64**</td>
<td>.64**</td>
<td>.60**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listens to others point of view</td>
<td>6</td>
<td>.49**</td>
<td>.64**</td>
<td>.80**</td>
<td>.48**</td>
<td>.61**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acts friendly toward others</td>
<td>7</td>
<td>.53**</td>
<td>.75**</td>
<td>.58**</td>
<td>.72**</td>
<td>.69**</td>
<td>.60**</td>
<td></td>
</tr>
</tbody>
</table>

** p < .01

Morning observation verses afternoon observations

The Pearson’s correlation was also used to evaluate the connection between play observation sessions (morning versus afternoon). The mean value for each child rated with the Prosocial/Communication Skills Scale was calculated across raters for each observation session. Excluding missing data a total of 109 children participated in both observations. The correlation between morning observations (M=3.50; SD=.95) and afternoon observations (M=3.88; SD=1.10) rated with Prosocial/Communication Skills Scale showed statistical positive significance (r=.37, p < .01 (2-tailed).

Frequency and rating patterns between raters of “did not observe”

Patterns between observers were found by comparing each case between raters and code them as presented in table 5. Each code is presented in table text. From table 5 it is clear than observers more frequently used “did not observe” during the morning observation than during the afternoon observation.
Table 5. Frequency over how ratings between observers

<table>
<thead>
<tr>
<th>Observation 1 = morning</th>
<th>1* = Total agreement</th>
<th>2* = Close agreement</th>
<th>3* = Agreement “did not observe”</th>
<th>4* = Difference 2 values or more</th>
<th>5* = One using 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial/communication Scale Observation 1</td>
<td>20</td>
<td>22</td>
<td>24</td>
<td>66</td>
<td>12</td>
</tr>
<tr>
<td>Prosocial/communication Scale Observation 2</td>
<td>31</td>
<td>34</td>
<td>7</td>
<td>72</td>
<td>11</td>
</tr>
<tr>
<td>Emotional regulation Skills Scale Observation 1</td>
<td>27</td>
<td>17</td>
<td>24</td>
<td>68</td>
<td>5</td>
</tr>
<tr>
<td>Emotional regulation Skills Scale Observation 2</td>
<td>34</td>
<td>27</td>
<td>5</td>
<td>66</td>
<td>11</td>
</tr>
</tbody>
</table>

1* = Percentage when observers had a total agreement in ratings. 2* = Percentage of when there was a close agreement with only one value difference. 3* = Rating percentage when both observers rated 0 as in “did not observe”. 4* = Rating percentage when observers rated with two or more values difference. 5* = Rating percentage of when one used 0 as in “did not observe”. **Σ = Percentage of 1*, 2*, 3* combined, showing close and total agreement together.

In order to get a clear picture of the distribution of ratings Figure 1 and 2 was created (See appendix A and B) showing percentages over how observers rated on item level for each observation session and for each subscale. White and white/dot was used to show absolute agreement and close agreement (when observers only differed on one value). Gray areas and gray/stripe in the picture shows frequency over when both or one observer used “did not observe”. Frequencies when one observer used “did not observe” was also divided into two groups depending on how the other observer rated. This was to see if using “did not observe” was related to higher or lower ratings. Black was used to show then observers disagreed on more than 2 values. The main findings from these figures were the high percentage of observers using “did not observe”.

For the Prosocial/communication Skills Scale (SCO), Item 1 “Resolves peer problems on own” had a frequency of 84% where one or both observers used “did not observe” during the morning observation. On item 2 “Very good at understanding feelings” 75% consisted on one or both observers using “did not observe”. This high percentage and was consisting during both morning and afternoon observations (See appendix A for more details regarding all items).

For the Emotional Regulation Skills Scale (SCO) item 9: “Copes well with failure” had a frequency of 93 % where one or both observers used “did not observe” during the morning observation whereas this decreased to 31 % during the afternoon. Item 8 “Acts friendly toward others” also had a high frequently (84 %) when one or both used “did not observe” during the morning but the percentage decreased to 26 % during the afternoon. The rest of the items were rated with consistent patterns across items during the morning and afternoon observations with an average of high frequency of observers using “did not observe” (See appendix B for more details regarding all items).

From figures presented in appendix A and B it was also seen that when one observer rated “did not observe” the other observer more frequently rated the same item “moderately well” or higher instead of “not at all” or “a little”.

14
Discussion

The aim of this study was to evaluate some of the psychometric properties of the observer rating scale SCO in relation to a teacher rated scale of the same construct. More specifically, the focus of the thesis is on two sub-scales; Prosocial/Communication Skills Scale and the Emotional Regulation Skills Scale, both included in SCO and used in the Swedish adaptation trial of PsPATHS. With this study, new, opposing results in relation to previous reliability measurements of the instrument is provided. Overall, findings indicate that SCO or the method used when rating with SCO, has reliability flaws and is sensitive to this new context. Using unreliable measurement is a serious problem (De Vellis, 2003). Adding the high frequency of “did not observe” incorporated as missing data, SCO and the play observation methods, should within the Swedish Trial of PsPATHS be handled with restricted care for future use and data interpretation. It is also important to see if these results replicate with the entire PsPATHS data set. The findings further emphasize the necessity of critical evaluation of measurements and the need for several evaluations across different contexts to establish adequate reliability and validity (Cordier et al., 2016).

Reliability

ICC was calculated in order to establish the inter rater reliability between raters. For the Prosocial/Communication Skills Scale it was measured as “good”. For the Emotional regulation Skills Scale the interrater reliability was instead found to be “poor”. Compared to measurements conducted within the same research project but with another data set, Emotional regulation Skills Scale was measured to a value “excellent” [ICC] = .74 (Gulliksson, 2014). Results in this study then show a problematic decrease. Looking at the full scale level of SCO, previous studies has reported a “good” value, [ICC] = .70 (Bierman et al., 2008). The criterions of values are according to Cicchetti’s guidelines (1994). (CIs were not reported for these previous results). The low ICC for the emotional regulation scale indicates that items on that scale may be unclear or that observers have different perceptions regarding displayed behaviors related to each item. And although new observers got the chance to practice ratings and discuss them with supervisors in line with Smith’s (1986) guidelines to reach a high inter rater reliability, the ICC value further indicates that the training before the actual observations at the preschools could be improved and clarified.

Before comparing SCT with SCO, some reliability checks were conducted for SCT showing normal distribution, excellent value of internal consistency reliability using Cronbach α and high inter-item correlations between all items (p > .01) presented in table 1 and 3. The same procedure was done for SCO with the exception of calculation ICC instead of Cronbach α. Correlations between items shown in table 4 across raters and observation sessions were statistically significant (table 2). Based on these reliability checks and previous results of sound psychrometric properties of SCT, (Bierman et al., 2008) a high correlation between SCT and SCO would have indicated solid concurrent validity of SCO. Both scales were used with the same sample during the same time period according to guidelines when measuring concurrent validity (Carmines & Zeller, 1979).

Comparing Social Competence Scale –Teacher (SCT) and Observer (SCO)

A positive statistical significance was found in the comparison between SCO and SCT using the Prosocial/communication Skills Scale. However it was not strong enough to draw conclusions of concurrent validity. It is possible that the correlation might have
been higher if SCO had not consisted of such a high amount of “did not observe” but the result is very similar to Bierman’s Study (2008) who with 356 children, also got a weak positive statistical significant correlation between SCO and SCT ($r = 0.29, p < 0.05$). Although, Bierman et al (2008) were able to compare the full scale since ICC values were acceptable for the Emotional regulation Skills Scale as well as for the Prosocial/communication Skills Scale. Even though being significant, the correlations indicate that teachers and observes seems to have a different comprehension as to how to rate the children. As stated in the literature, measures of social competence have shown large variations, both between measurement methods and between assessors using the same instrument (Cordier et al., 2016; Denham, 2006; Rose-Krasnor, 1997).

In aspect of bias, both teachers and observers may have suffered difficulties in being subjective (Cordier et al., 2016). Comparing methods, the teachers did not have the “did not observe” option since they saw the children every day and therefore had the opportunity for richer data compared with observers who only met the children during two fifteen minutes short intervals. Looking at the literature, scholars have highlighted the benefits by observing children in their natural setting and have found them equally valid to more traditional standardized methods (Denham, 2006, Farmer-Dougan and Kaszuba, 1999). The strength of relationship has also been regarded as an important factor concerning children’s tendency to show prosocial skills (Eisenberg, 1983). The play observations would then indicate that the presence of strangers might have affected how the children behaved, and the group setting itself. If the group of children usually did not play together, the setting could have been preserved as unnatural and strange. Future researcher might therefore consider using more naturalistic setting in order to capture social and emotional competence. Both because of richer data but also to be able to assess the child’s true behaviors based on ecological validity.

“Did not observe” and Sensitivity to context

One obvious change in the scale compared with earlier measurements within the research project of PsPATHS in Sweden is the addition of the rating option “did not observe”. During the pilot studies A and B items that had not been observed were left blank. However, by doing this is it was unclear whether the observer had not observed the behavior or if it was just not recorded. Regardless of method to rate, “did not observe” ratings were later viewed as missing data. Though, looking at Table 5 it is clear that the frequency to choose “did not observe” overall all made up at least one quarter of the data-set. During the morning play observation it covers 46 % of the data on the Prosocial/communication Skills Scale and 51 % on the Emotional regulation Skills Scale (Including both when one or both observers used “did not observe”). Missing data with this high frequency has been regarded to be very problematic in terms of future conclusions (Howell, 2012). With these findings, it is reasonable to question whether observers used the alternative when they hesitated or if these kinds of situations actually did not occur. Looking at the greater percentage of the use of “did not observe” in the morning in comparison to the afternoon, the latter explanation might be more reasonable. If observers used “did not observe” due to hesitation the scoring would likely have been equally distributed between observation sessions.

The statistically significant correlation between morning and afternoon observations points out that the observation sessions are roughly similar. However, as stated above and with regard to that the procedure should be performed in the exact same way, with the same raters and the same group of children with the only exception to use different
toys, it would have been expected to get possibly a higher correlation between ratings on the sessions. In a previous study with the play observation sessions conducted within the same research project, but on the pilot data, the correlation was higher \( r = .53, p < .01 \) (Gulliksson, 2014).

Although the morning and afternoon sessions are similar, the differences of using “did not observe” (table 5) between the scales and sessions, suggests that the context cannot be completely discounted. Instead the setting seemed to have some impact on how the observers rated the children or the children’s actual behaviors. During the morning play observations observers and children interacted for the first time. As a consequence, children tended to sometimes be more shy and reserved. During the afternoon, the child had participated in the morning play observation as well as individual testing were usually more relaxed and excited. Additionally the children had experience of the group setting and who they were going to play with during the afternoon because the groups were fixed or the same in the morning and afternoon. Additionally, observers were also more familiar with the children and might have been more able to observe their behaviors. The toy itself could also have affecting the moods of children. The pony farm may have resulted in a higher degree of calm play whereas the marble run which with very few exceptions, tended to collapse during the play. Usually the collapse resulted in more interactions than the pony farm.

On an item level it is clear that some items less frequently occur or are regarded as more complex to observe. Figure 1 and 2, together with previous studies, (Kvamme, 2014) shows that item 1 “Resolves peer problems on own” and item 9 “Copes well with failure consistently have resulted in low occurrence of ratings. A possible explanation for the low occurrence is that failure and peer problems had to occur in order to cope and solve them. During friction free play, problems might not arise. In this sample, these items as well as item 2 “Very good at understanding feelings” were difficult to observe across both morning and afternoon observations. For future use of SCO these are recommended to be reworded, replaced or excluded.

**Limitations**

There are many possible limitations in the present study. In both studies, observers were allowed to take narrative notes during the whole play observation, trying to capture all three children at the same time. Based on these notes, observers were free to rate the children using SCO which might have led observers to have different perspectives regarding observed behaviors. According to Miller, (2007) this is an example of narrative recoding in the least structured form in risk of bias. During these free setting children who showed a lot of behaviors might also have get more attention from observers than children showing more subtle behaviors. The method has been hard to control and it is impossible to say that observers perceive the same behaviors, of the same child, during the same time. While writing narratives, one observer could easily have missed behaviors that another one recoded. During the last evaluation (Gulliksson, 2014), a recommendation was to use a narrower and more controlled structure in which each child was observed individually, dividing the play observation between children. However, with the high rate of “did not observe” this method would stand the risk of even more missing data. In addition, both observations were conducted during the same day which is a problem in terms of reliability. Furthermore, one observer had previous experiences of the scale and assessment and was educated at a different time with different instructors which might have caused mixed messages regarding how the ratings should be conducted.
Competencies in prosocial behaviors such as sharing and helping, together with emotional competence like emotion regulation, has shown be important in order to establish and maintain meaningful relationships (Bierman et al., 2008, Denham, 2007; Domitrovich et al., 2007) Considering these positive outcomes as well as the resources and time consumption large scale interventions such as PsPATHS implies, it is crucial to use measurements that results in high quality and quantity responsiveness (Denham, 2006). This thesis confirms the need to gain insight on appropriate measurement methods on how to capture social and emotional competencies among young children stated by Zaslow et al., (2006). The measured psychometrical properties and the review of “not observed” behaviors can be of value for future use and crucial to consider within the Swedish trial of PsPATHS. This thesis has provided evidence on the importance of repeated measurements in order to establish a high validity and reliability.

References


Appendix A. Figure 1. Frequencies in percentage over of how observers rated per item and observation session for Prosocial/communication Skills Scale

Item index: Pro Social Communication Skills Scale
1. Can solve peer problems on own
2. Very good at understanding feelings
3. Shares material
4. Cooperate with friends without prompting
5. Is helpful to others
6. Listens to others point of view
7. Acts friendly toward others

Pro Social Communication Skills Scale, SCO

Morning observation
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7

Afternoon observation
Item 1
Item 2
Item 3
Item 4
Item 5
Item 6
Item 7

- Absolute agreement
- Close agreement
- Total agreement of "did not observe"
- Rater 1= 0 Rater 2= >2
- Rater 1= 0 Rater 2= <2
- No agreement
Appendix B. Figure 2. Frequencies in percentage over of how observers rated per item and observation session for Emotional Regulation Scale

Item Index; Emotion Regulation Skills Scale
8. Accept things not going well
9. Copes well with failure
10. Accept legitimate limits
11. Expresses needs and feelings
12. Think before acting
13. Aware of effects of own behaviors
14. Plays by rules of game

![Graph showing frequencies in percentage over of how observers rated per item and observation session for Emotional Regulation Scale.](image)
Appendix C. Used Items of the Social Competence Scale, Observer Version.

Participant Name or Number: _______________________
Preschool: ______________________________________
Date of Observation: ______________________________
Observer Name: __________________________________

<table>
<thead>
<tr>
<th>Prosocial/Communication Skills</th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately well</th>
<th>Well</th>
<th>Very well</th>
<th>Did not Observe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Resolves Peer Problems on Own</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>2. Very Good at Understanding Feelings</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>3. Shares Materials</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>4. Cooperates with Peers without Prompting</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>5. Is Helpful to Others</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6. Listens to Other Points of View</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>7. Acts Friendly Toward Others</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Emotional Regulation Skills</th>
<th>Not at all</th>
<th>A little</th>
<th>Moderately well</th>
<th>Well</th>
<th>Very well</th>
<th>Did not Observe</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. Accepts Things Not Going Way</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>9. Copes Well Failure</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>10. Accepts Legitimate Limits</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>11. Expresses Needs and Feelings</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>12. Thinks Before Acting</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>13. Aware of Effects of Own Behavior</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>14. Plays by Rules of Game</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
Appendix D: Observation Code Book: Developed within the project of the Swedish Adaptation of PsPATHS.

Prosocial Communication Skills Scale

1. Resolves peer problems on own
   (i.e. does not turn to a teacher or observer for help in a peer conflict; approaches peer with a calm demeanor and uses words/language to come to a resolution)
2. Very good at understanding feelings
   (e.g. pays attention to other children and in the case of another child’s conflict/frustration or other emotional display, responds accordingly)
3. Shares materials
   (i.e. no ‘snatching’; ‘hoarding’ toys, materials with body or language)
4. Cooperates with peers without prompting
   (e.g. works together with peers; collaborates on a task or problem with a peer or assists in a task with peers)
5. Is helpful to others
   (e.g. assists, offers or is ready to assist peers – i.e. if a marble falls of the table, the child helps their peer to fetch it)
6. Listens to other points of view
   (e.g. hears and listens to peers – not necessarily compliant, but displays positive communication and/or exchange)
7. Acts friendly towards others
   (e.g. the child exhibits kindness towards peers and is pleasant – i.e. a child is not just helpful or cooperative during task execution, but displays an explicit, outward and deliberate ‘friendly demeanor’/a kind exchange/affection – i.e. a smile or ‘little chat’/communication with a peer)

Emotional Regulation Skills

8. Accepts things not going his/her way
   (i.e. following a conflict with a peer, the child does not keep a toy - e.g. does not have an emotional outburst, rather attempts to problem solve, uses words, plays positively following event, etc.)
9. Copes well with failure
   (i.e. attempting to fix or put together an aspect of the toy – ‘just doesn’t fit or work’ - e.g. does not have an emotional outburst, rather attempts to problem solve, uses words, etc.)
10. Accepts legitimate limits
    (e.g. playing within the ‘play-limits’ – i.e. physical space and time limitations, limits of the toy, play situation, being with peers and sharing, etc – without negative expression of emotion/being positive throughout the play)
11. Expresses needs and feelings
    (e.g. without a heightened arousal, being able to express him/herself, especially in a conflict/upsetting situation)
12. Thinks before acting
    (e.g. explicit control of impulsive behavior – i.e. child plays in an intentional and thoughtful manner) ***does the intent matter? i.e. child thinks and intentional and hurts another child?
13. Aware of effects of own behavior
    (e.g. child displays awareness of their actions and how they affect peers) ***does the intent matter? i.e. child thinks and intentional and hurts another child? – does this include empathy
14. Plays by rules of the game*
    (e.g. refers to explicit and implicit rules of play – social norms, interaction with peers and with toy and space – i.e. does not take the toy out of the room, or the main toy)

*Explicit rules of the game refer to “Play here, with this toy, together, for 20 minutes”
(observer says this at the beginning of the play)