The Icelandic translation of the Strengths and Difficulties Questionnaire - Teacher Form.

An investigation of the psychometric properties in a sample of 6-10 year old children.

Einar Kári Bogason

Lokaverkefni til BS-gráðu
Sálfræðideild
Heilbrigðisvísindasvið

HÁSKÓLI ÍSLANDS

The Icelandic Translation of the Strength and Difficulties Questionnaire - Teacher Form

An investigation of the psychometric properties in a sample of 6-10 year old children.

Einar Kári Bogason Lokaverkefni til BS gráðu í Sálfræði

Lokaverkefni til BS gráðu
Leiðbeinandi: Urður Njarðvík Ph.D og Fanney Þórsdóttir Ph. D.
Sálfræðideild
Heilbrigðisvísindasvið Háskóla Íslands
Júní 2011

Prentun: Prentsmiðja: Háskólaprent.
Staður, Ísland 2011

Table of contents

Objective: To examine the psychometric properties of the current Icelandic translation of the Strength and Difficulties Questionnaire (SDQ) teacher form in a community sample of 6-10 year old children.

Method: 16 teachers rated 5 children in their classes using the current Icelandic SDQ totalling in 80 ratings in all. The children were chosen at random by the experimenter. Two factor analyses were conducted, one in which no specific number of factors were to be drawn and one in which five factors were determined to be drawn. This decision was based on the original five factor structure of the SDQ and the fact that the first scree plot seemed to indicate a five factor structure. In both cases a principal component analysis with a varimax rotation was used.

Results: The original five factor structure was not replicated in either factor analyses. The non-predetermined factor structure analysis produced 8 factors in total with Eigenvalues over 1. Both factor analyses reproduced the hyperactivity scale, though it was not as clean in the five factor structure analysis. The first factor analysis also reproduced the peer problems subscale, which the second factor analysis failed to do. The second factor analysis did however reproduce the prosocial subscale, which the first one failed to do.

Discussion: Some items proved quite problematic in this limited sample, though it seemed that some of those were not caused by the small sample size. Those items may need revision, but nothing conclusive can be drawn from such a small sample. Goodmans (1997) suggestion to use the 90th percentile as a clinical cut-off in each subscale did not seem appropriate in this study since the 90th percentile often landed on scores that were deemed to be in the normal range in UK normative data. Once again, nothing conclusive can be drawn from this due to the small sample size.

Conclusions: The results from this study indicates that the current Icelandic translation of this questionnaire is not suitable for use in this age group. It did not replicate the original five factor structure, nor did it provide a useful alternative factor structure. It is not clear whether this is due to the small sample size or to peculiarities in the items themselves. Further study is needed.

We tend to think of childhood and adolescence as the care-free days of our lives. A time when we were marvellously free of any and all troubles that only befall us in adulthood. This might actually be true in some cases, but it is not so for a lot of children. Contrary to popular belief, children can suffer from a myriad of mental and behavioural problems. Some of these might be attributed to situational causes, such as lack of sleep or consuming too much sugar. Although sleep deprivation can make us a bit cranky, it is far from always the cause of symptoms resembling mood disorders, and sugar does not cause ADHD.

Even in the scientific literature, little consideration was given to the well-being of children until the 18th century, and children were considered the exclusive property of their parents (Mash & Wolfe, 2010). However, inspired by philosophers such as John Locke, who argued that children are born as more or less blank slates (tabula rasa), the idea that we should raise our children to become productive members of society instead of merely punishing them for being driven by primal urges, began to take hold in the scientific community as well as society at large (Mash & Wolfe, 2010). Although, no real societal concern for the well being of children awoke until the 1800's, when inquiries about how long children could work in factories and mines were conducted in England. Though the study was mainly conducted out of practical concerns (Cole, Cole & Lightfoot, 2005).

In those days gone by, many in the scientific community believed that children could not suffer from psychological problems. For instance, Freud claimed that children could not become depressed since they did not have a fully developed Superego (Mash & Wolfe, 2010). Whether or not there is such a thing as a superego and regardless of how well developed such a thing would be in children, we know that children can become depressed. However, in his theories, Freud claimed that adult psychopathologies were rooted in childhood experiences (Mash & Wolfe, 2010). So to quite some extent, Freud's theories on psychopathology were connected to child development. He even developed an extensive framework for psychosexual development of children, disturbances of which, he claimed, caused pathology which became apparent in adulthood (Ross, 1987). So a fixation in a specific stage of psychosexual development could produce a specific type of symptoms in adulthood. For example, a fixation in the anal stage could manifest itself as obsessive-compulsive disorder and a narcissistic personality disorder could arise

if a fixation occurred in the narcissistic phase, before love for themselves was shifted onto significant others (Noelen-Hoeksema, 2008). He had thus set in motion the view that when it comes to psychological difficulties with children, it is exceedingly important to take into account that they, unlike adults, are still developing. This is a key factor when dealing with psychological difficulties in children.

This fact is a central point in developmental psychopathology, which is a discipline in which abnormal behaviour is assessed with respect to developmental milestones children of specific ages usually have reached (Mash & Wolfe, 2010). It also tries to determine when a behaviour is a normal, transient reaction to childhood difficulties and when it turns into a maladaptive response. (Noelen-Hoeksema, 2008) For instance it is entirely normal for 3-month olds not to speak words, but it's not so for 2 year-olds. This is a very simplistic example but it does reflect the intervals we must deal with. The young brain is still developing and individual differences are vast. Still, given individual differences and plasticity, the ages of which these milestones "should" be reached would be better considered guidelines than developmental absolutes. The fact that the field of developmental psychopathology exists reflects the importance of accounting for the state of ever changing development the children are in when assessing psychopathology in them.

Since children are still developing, interventions are not aimed at returning the patient/client to a previous state of functioning, but rather to boost their existing abilities (Mash & Wolfe, 2010). This is most effectively done by using their strengths to build up weaker traits to make them more well-rounded individuals. It is thus very important to assess both the strengths *and* difficulties of a child.

An instrument that assesses both strengths and difficulties would thus prove immensely useful. One such tool is the Strengths and Difficulties Questionnaire (SDQ) developed by Robert Goodman (Goodman, 1997). This study will evaluate the latest Icelandic translation of the teacher-form SDQ for children aged 6-10.

Mental disorders and prevalence in children and adolescents.

There are several disorders that are commonly first diagnosed in childhood and adolescence. The most commonly and severe diagnosed disorders are behavioural disorders, mood disorders, anxiety disorders and developmental disorders (Mash & Wolfe, 2010).

Behavioural disorders

The three most common behavioural disorders are Attention Deficit/Hyperactivity disorder (ADHD), Conduct Disorder (CD) and Oppositional-Defiant disorder (ODD).

ADHD is characterized by (as the name implies) lacking attention and hyperactivity which are age-inappropriate and cause impairment for the child and others.

Prevalence rates vary from about 3-7% in school-age children when the DSM-IV-TR criteria are used (American Psychiatric Association (APA), 2000). The diagnostic criteria of the DSM-IV state that the symptoms must be present for at least 6 months, some symptoms must have been present before the age of 7 and there must be impairment in at least two settings (APA, 2000). The symptoms are split into two distinct categories and in order to be diagnosed the child must display at least 6 of those symptoms. The reason for the split is that the DSM-IV makes a distinction between different types of ADHD. There is the Predominantly inattentive type (ADHD-PI), which is mostly characterized by symptoms of inattention, the Predominantly hyperactive-impulsive type (ADHD-HI), which is characterized by primarily hyperactive and impulsive symptoms, and the Combined type (ADHD-C), which have symptoms of ADHD but cannot easily be placed in any of the other two categories (APA, 2000).

Conduct disorder (CD) is characterized by "*a repetitive and persistent pattern of behavior in which the basic rights of others or major age-appropriate societal norms or rues are violated*" -APA (2000)

The 15 behaviours listed by the DSM-IV are grouped into four broader categories. These are:

- Aggressive conduct that causes or threatens physical harm to other people or animals

- Non-aggressive conduct that causes property loss or damage
- Deceitfulness or theft
- Serious violations of rules

In order to meet the diagnostic criteria the individual must display at least three of the 15 behaviours in the past 12 months with at least one behavioural criteria present for at least 6 months.

Prevalence rates range from 1-10% in general population studies and are more common amongst males than females (DSM-IV, 1994), which is hardly surprising since it is defined in externalizing terms which are much more commonly found in males than females of this age (Bongers, et. al., 2003) CD might be a precursor to adult antisocial personality disorder, since about 40% of children diagnosed with CD develop antisocial personality disorder as young adults (see Mash & Wolfe, 2010).

Oppositional defiant disorder (ODD) is a constellation of age-inappropriate behaviours of negative, disobedient, defiant and hostile behaviour towards parents and other authority figures that persist for at least 6 months (APA, 2000). It usually has its onset around age 8 and may be a precursor to CD or a milder form of it since there is much overlap in symptoms and approximately 25% of children diagnosed with ODD are later diagnosed with CD (see Mash & Wolfe, 2010). Prevalence rates of 2-16% have been reported in various populations and by various methodologies (APA, 2000).

Mood disorders

Even though there are no mood disorders that are specific to childhood, it is important to highlight a few here since they can impair development. It's very difficult for children and youths to focus on school-work and their social development when they're depressed and this can have serious repercussions in adulthood.

Depression is a very common disorder, and approximately 20% of children and adolescents experience depression sometime in their life (Avenevoli, Knight, Kessler & Merikangas, 2008). Isolated incidences of depressive symptoms are usually temporary, but if they persist and are severe enough to cause impairment, then the individual may be diagnosed with Major Depressive Disorder (MDD). The symptoms of MDD include depressed mood, loss of interest in activities that were

previously enjoyed, hopelessness and sleep disturbances (Mash & Wolfe, 2010) and they must persist for at least two weeks (APA, 2000). Adolescents are more prone to depression than children, with prevalence rates of about 8% in adolescents, 2% in school-aged children and less than 1% in pre-school children. In younger age-groups there is little gender difference in prevalence rates, however a marked difference occurs at around 15 years of age and at around 18 years of age females are twice to three times likelier to be diagnosed with MDD (see Mash & Wolfe, 2010).

Dysthimic disorder is related to MDD and is sometimes considered a subcategory of depression. It is longer-lasting than depression but the symptoms are less severe. Prevalence rates range from 1% in children to 5% in adolescents (see Mash & Wolfe, 2010) and usually has an earlier onset than MDD, in fact approximately 10% of those diagnosed with Dysthimic disorder (in epidemiological samples, 15-25% in clinical samples) also develop MDD (APA, 2000).

Anxiety Disorders

Children can become affected by many anxiety disorders. But one anxiety disorder, separation anxiety disorder, is specific to children (Noelen-Hoeksema, 2008).

Separation anxiety disorder is characterized by unusual amounts of distress when separated from or when anticipating desperation from caregivers or their home and persistent worries about losing or harm coming to their caregivers (APA, 2000). The consequences of these worries are reluctance or refusal to leave the home or their caregivers. This has potentially adverse consequences on a child's social and academic development, since they avoid going to school and social activities which may take them away from their caregivers. Prevalence rates average around 4% and decrease in adolescence. It is more common in girls than boys in epidemiological samples, though roughly equal in clinical samples (APA, 2000). Clinical referrals have been made for children as young as 7-8 years of age (see Mash & Wolfe, 2010).

Assessment of children

There are numerous ways of assessing psychological difficulties in children. The method of assessment most commonly associated with psychological assessment is the clinical interview. These of course provide a lot of information about one individual child, but they're very time-consuming and can also be quite expensive. This method of assessment is thus not a very reasonable tool to use when doing larger community prevalence studies or screenings.

A useful way of conducting such studies is via multi-informant screening questionnaires. These questionnaires are filled out by parents, teachers or the children/adolescents themselves. A screening questionnaire will aim to identify individuals who may have, or be at risk of developing, psychological difficulties which can then be referred to further evaluation. Since each type of rater has different experiences and perceptions of the child, it provides important information about when and where the behaviours occur (Smith, 2007). If a child displays problematic behaviours in school but not at home or in social activities then the problem could be a symptom of school-related difficulties (such as learning- or reading difficulties), rather than some other disorder. These questionnaires also provide information about the perceptions of the raters. If one parent rates the child very negatively but the other parent rates it much more positively, then we must examine whether this difference is due to genuine differences in experience or if the difference lies in perception. For instance one of the parents might have psychological difficulties of his/her own or has had little experience with the child, which might lead to distorted or inadequate perceptions (Smith, 2007). Such contrasting information would not be available if only one rater was used or in a single-subject clinical interview. The researcher or clinician would have to rely on the reliability of the information given to him/her by the single rater or interviewee.

Behavioural screening questionnaires do not measure the actual occurrence of difficulties but rather the perceived difficulties as they are experienced by either an external rater (e.g. parents or teachers) or by the child itself (Smith, 2007). None of these raters are likely to give complete information about the behaviours since they often vary between situations (Achenbach, 2001).Thus it makes good sense to

incorporate as many raters as possible, so one can obtain as diverse information as possible and compare the ratings.

Another advantage screening questionnaires have over clinical interviews is that they make comparisons easier. In order to ensure complete comparability with interviews, a series of interviews must be conducted in exactly the same way and scoring must be precise and consistent. This is not an unaccomplishable goal, of course, but it's much easier to ensure consistency with a questionnaire. A well developed questionnaire will be highly standardized and scores will be easier to make sense of, since it will compare the score of the child with a known reference group.

About the Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a behavioural screening questionnaire developed in England by Robert Goodman which is intended to assess difficulties in youths and children of 4-16 years of age (Goodman, 1997). There are parent and teacher forms for the completion on 4-16 year-old children and youths and self-report forms for 11-16 year-old youths (http://sdqinfo.org). There are also parent- and teacher forms for 3-4 year-olds where items that assess antisocial behaviour are switched for items that assess oppositionality (National Child Traumatic Stress Network [NCTS]), since this is more age-appropriate. It has since been translated to over 40 languages (NCTSN.org), amongst which most of the Nordic languages (Obel, et al, 2004).

The SDQ was designed based on a factor analysis conducted by Goodman in 1994 (see Goodman, 1997) on a modified version of the parent form Rutter questionnaire. The modification consisted of the inclusion of many, mainly positive items. The original 31 items of the Rutter parent form was coupled with 26 items from the Rutter teacher questionnaire, along with 4 items from the parent form and 20 items from the teacher form of the Prosocial Behaviour Questionnaire (see Agnes Huld Hrafnsdóttir, 2005). The factor analysis suggested that the questionnaire was assessing five distinct dimensions, which are now used in the modern SDQ (Goodman, 1997). The five dimensions are: conduct problems, emotional problems, hyperactivity, peer problems and prosocial behaviour. Each of these five dimensions are assessed by five items, making the list consist of 25 items total. This makes the

SDQ quite a short screening questionnaire, especially when compared to the 118-item long Child Behavior Checklist [CBCL].

As the name would imply, the SDQ assesses not only difficulties or weaknesses, but also strengths. It does so by assessing prosocial behaviour and by having positively worded items, which are scored in reverse. So the items *"Thinks things out before acting"*, *"Sees tasks through to the end, good attention span"*, "*Has at least one good friend*", "*Generally liked by other children"* and "*Generally obedient, usually does what adults request"* are in fact (for clinical purposes anyway) assessing inattention, hyperactivity, peer problems and conduct problems, since they are negatively scored (Goodman, 1997). These five positively worded items plus the five items assessing prosocial behaviour constitute the list's 10 "strength items". There are 14 questions assessing difficulties and one item "Gets on better with adults than with other children" is considered neutral (Goodman, 1997). The questionnaire was designed to include positive items and be short to make it more acceptable to respondents (Goodman & Scott, 1999). This would be particularly relevant when using a community sample, as they might not be as motivated as a clinical sample to complete a long and exclusively negative questionnaire.

Additionally, because of it's brevity it enables researchers to identify more children with difficulties in a shorter amount of time, thus reducing costs. Furthermore, there is no training required to administer the test, nor do you need to pay any licensing fees to use the SDQ as long as you yourself do not charge anyone for your services in connection to the usage of the SDQ.

The SDQ was designed to be a behavioural screening questionnaire with these distinct criteria: It should be no longer than 1 page (A4), it should cover the ages of 4-16 years, the teacher and parent forms should be identical and a similar version should exist for self-report (suitable for 11-16 year-olds), both strengths and difficulties should be assessed and there should be an equal number of items tapping each of the five dimensions (Goodman, 1997).

The Rutter questionnaires the SDQ is built on are about 40 years old and are thus fairly dated. They inquire solely about difficulties and do little to assess a child's strengths, which is contrary to recent trends, which has been to emphasise both strengths and weaknesses (Goodman, 1997). In fact, it is very useful to identify a child's strengths, since these may serve as protective factors and a starting point

from which to guide the child to a more adaptive life (Mash & Wolfe, 2010). Additionally, the Rutter questionnaires include items which seem to have little conceptual value, such as nail-biting and thumb sucking, whereas conceptually valid dimensions, such as hyperactivity, inattention and sociability have little or no coverage. The SDQ on the other hand, was specifically designed to contain items that tap nosological concepts that underpin ICD-10 and DSM-IV classifications (Goodman & Scott, 1999). For example, the hyperactivity/inattention scale contains two items that assess hyperactivity, two which tap attention and one that inquires about impulsivity, which are all key symptoms for a DSM-IV ADHD diagnosis (the DSM-IV diagnostic criteria for ADHD has six symptoms for hyperactivity, six for inattention and three for impulsivity (APA, 2000).

Subscales

Robert Goodman describes the SDQ in detail in his initial study of the SDQ as such (Goodman, 1997). The SDQ consists of five subscales, each containing five items, totalling in 25 items.

These scales are: emotional symptoms, hyperactivity, conduct problems, peer problems and a prosocial scale. Each scale in the English SDQ contains the following items:

Emotional Symptoms Scale: “Often complains of head-aches, stomach-ache or sickness”; “Many worries, often seems worried”; “Often unhappy, down-hearted or tearful”; Nervous or clingy in new situations, easily loses confidence” and “Many fears, easily distracted”.

Hyperactivity Scale: “Restless, overactive, cannot stay still for long”; “Constantly fidgeting or squirming”; “Easily distracted, concentration wanders”; “Thinks things out before acting” and “Sees tasks through to the end, good attention span”

Conduct Problems Scale: “Often has temper tantrums or hot tempers”; “Generally obedient, usually does what adults request”; “Often fights with other children or bullies them”; “Often lies or cheats” and “Steals from home, school or elsewhere”.

Peer Problems Scale: “Rather solitary, tends to play alone”; “Has at least one good friend”; “Generally well liked by other children”; “Picked on or bullied by other children” and “Gets on better with adults than with other children”

Prosocial Scale: “Considerate of other people's feelings”; “Shares readily with other children (treats, toys, pencils, etc.)”; “Helpful if someone is hurt, upset or feeling ill”; “Kind to younger children” and “Often volunteers to help others (parents, teachers, other children)”.

Scoring the subscales

Each item is scored on a 3-point likert scale, with the options: “Not true”, “Somewhat true” and “Certainly true”. 20 of 25 items are scored 0 for “Not true”, 1 for “Somewhat true” and 2 for “Certainly true”. The remaining five items: “Thinks things out before acting” (Hyperactivity scale), “Sees tasks through to the end, good attention span” (Hyperactivity scale), “Generally obedient, usually does what adults request” (Conduct problems scale), “Has at least one good friend” (Peer problems scale) and “Generally liked by other children” (Peer problems scale) are scored in reverse. Meaning that “Not true” is scored as 2, “Somewhat true” is scored as 1 and “Certainly true” is scored as 0. The assumption behind the reverse scoring is that the absence of that which the items inquire indicates difficulties (Goodman, 1997).

Each scale contains 5 items, each of which can be given a score ranging from 0-2, which means that each scale will have a score ranging from 0-10. The total difficulties score is computed by adding up the scores from 4 of the 5 subscales. Thus the total difficulties scale will have a score ranging from 0-40 with a higher number indicating a higher probability of pathology. The prosocial scale (which contains only positively worded items) is not reverse-scored and included in the total difficulties score since a lack of prosocial behaviours does not necessarily imply psychological difficulties (Goodman, 1997).

There are no official cut-off scores for the SDQ, but Goodman (1997) suggests that scores in the top 90th percentile should be considered to be in the abnormal range, the top 80th - 90th percentile in the borderline range and the bottom 80%

should be considered to be the normal range. In the same article Goodman (1997) supplies provisional bandings of SDQ scores which could serve as a useful guide when using the SDQ. These bandings are presented in table 1 below.

Table 1. UK bandings for the SDQ

	Normal	Borderline	Abnormal
Parent completed			
Total difficulties	0-13	14-16	17-40
Emotional Symptoms Score	0-3	4	5-10
Conduct Problems Score	0-2	3	4-10
Hyperactivity Score	0-5	6	7-10
Peer Problems Score	0-2	3	4-10
Prosocial Behaviour Score	6-10	5	0- 4
Teacher completed			
Total Difficulties Score	0-11	12-15	l6-40
Emotional Symptoms Score	0-4	5	6-10
Conduct Problems Score	0-2	3	4^10
Hyperactivity Score	0-5	6	7-10
Peer Problems Score	0-3	4	5-10
Prosocial Behaviour Score	6-10	5	0- 4

The subscales and items of the version of the Icelandic SDQ which is being tested in this study are:

Tilfinningaleg einkenni:

,,Er oft óhamingjusam/ur, langt niðri eða tárast", ,,Kvartar oft um höfuðverk, magaverk eða flökturleika", ,,áhyggjur á mörgu, virðist oft áhyggjufull/ur", óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust" og ,,óttast margt, verður auðveldlega hrædd/ur"

Hegðunarerfiðleikar:

,,Almennt hlýðinn, gerir yfirleitt eins og fullorðir óska", ,,Fær oft skapofsaköst eða er uppstökkur", ,,stelur heima, í skóla eða annars staðar", ,,lýgur oft eða svindlar" og ,,Flýgst oft á eða leggur börn í einelti"

Erfiðleikum í samskiptum við jafnaldra:

,,Frekar einræn/n, leikur sér oft ein/n", ,,Á að minnsta kosti einn góðan vin", ,,Öðrum börnum líkar almennt vel við hann/hana", ,,Verður fyrir stíðni eða einelti af hálfu annarra barna" og ,,Semur betur við fullorða en önnur börn".

Ofvirkni/athyglisbrest:

,,Eirðalaus, ofvirk/ur, getur ekki verið kyrr lengi", ,,Hugsar áður en hann/hún framkvæmir", ,,Fylgir verkefnum eftir til enda, heldur góðri athygli", Stöðugt á fíkt eða íði" og ,,Truflast auðveldlega, hann/hún á erfitt með að einbeita sér".

Félagshæfni:

,,Tekur tillit til tilfinninga annara", ,,Á auðvelt með að deila með öðrum börnum (nammi, dóti, blýöntun o.s.frv.)", ,,Hjálpsamur/söm ef einhver meiðir sig, er í uppnámi eða líður ílla", Góð/ur við yngri börn" og ,,Býðst oft til að hjálpa öðrum (Foreldrum, kennurum, öðrum börnum)"

Extended SDQ

Apart from the 25-item symptom scale, there is also an extension called the Impact supplement which consists of five items. It was designed to be completed on 4-16 year-olds by teachers and parents and there is a self-report supplement for 11-16 year-olds. The supplement asks the respondent if he/she believes that the child/youth in question has a problem. If answered in the affirmative the supplement inquires further about it's chronicity, the stress it causes the young person, the social impairment it may cause and the burden it causes others (such as it's family or school class). There are different versions available to all three rater-types covered by the SDQ. They differ in wording, such as that the parent version inquires about "*your child*", the self-report version asks about "*you"* and the teacher form asks about "t*his child*" or *"the child"*. In addition, the teacher addition does not inquire about whether the problem causes interference at home, in friendships or in leisure activities since it is not reasonable to assume that they would have access to this information.

The parent and self-report forms, however, ask about these three dimensions since they would have enough experience to make an assessment. Both the parent and the teacher forms ask about whether or not the problem interferes with the child's classroom learning, but the teacher form also asks about interference in peer relationships. The differences in content and wording reflect the importance of using multiple informants, as each class of informant has different experiences (Smith, 2007), and it's important not to ask questions that there is no reason to believe the rater would have access to, since this might cause the rater to lose faith in the validity of the test.

The usefulness of the impact supplement was demonstrated by Goodman (1999) when he found that the impact score (which consists of the distress & social impairment scores) was better able to discriminate between a clinical and a non-clinical sample than the symptom checklist. The question "Overall, do you think that this child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?" was almost as good a discriminator as the impact score.

The first question in the supplement (teacher form) asks about perceived difficulties:

"Overall, do you think that this child has difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?" and like the other questions in the supplement it is measured on a 4 point scale. Each of the four options in each of the five questions correspond to a number:

No = 0

Yes – minor difficulties = 1

Yes – definite difficulties = 2

Yes – severe difficulties = 3

If the first question is marked with "No", then the rater is asked not to answer the other four questions, since they measure the properties of the difficulty.

The second question asks about the chronicity of the problem:

"*How long have these difficulties been present?"*. With the options: *Less than a month; 1-5 months; 6-12 months;* and finally*: Over a year.* Which correspond to the numbers 0, 1, 2 & 3.

The third and fourth questions make up what is known as the impact score.

The third question asks about distress to the child:

"*Do the difficulties upset or distress the child?"* with the options: *Not at all* (0)*; Only a little* (1)*; Quite a lot* (2) and *A great deal* (3).

The forth question asks about social impairment caused by the difficulty. "*Do the difficulties interfere with the child's everyday life in the following areas?"*, with the options: *Not at all* (0)*; Only a little* (1)*; Quite a lot* (2) and *A great deal* (3).

For the parent and self-report forms, there are 5 items that make up the impact score. Four on social impairment and one on distress. This gives the impact score a range of 0-15 for these rating forms. The teacher impact score consists of only 3 items (2 on social impairment and 1 on distress), giving it a range of 0-9. Goodman (1999) suggests an alternative scoring system of the impact score, which is aimed to be more suitable to clinicians. In this scoring system, the options are scored: Not at all/Only a little = 0, Quite a lot = 2, and A great deal = 3. This makes

sure that a rating has to go above “Only a little” in order to rise above zero. With this scoring system the range of scores go from 0-15 for parents and self-report to 0-10 and from 0-9 for teachers to 0-6.

The fifth and final question, known as the burden rating, asks about the burden the child/youths problem puts on the family, class or if it makes life harder for the people around him/her.

In the teacher form the question is: “*Do the difficulties make it harder for those around you (family, friends, teachers, etc.)?*”. With the options: *Not at all (0), Only a little (1), Quite a lot (2)* and *A great deal (3)*

Scores from the impact supplement are usually treated as continuous variables to give researchers an idea of the dimensions of the difficulties. The impact score (distress and social impairment score) can be classified as normal, borderline or abnormal. An impact score of 0 is normal, 1 is borderline and 2 or more is considered abnormal (http://sdqinfo.org).

Psychometric properties of the SDQ

Factor analysis

A factor analysis is a way of reducing the amount of data by clustering them together based on correlations. When two or more items correlate highly with each other they can be treated as one variable or factor. This statistical technique is very useful when one is trying to assess the dimensions a questionnaire is measuring. This is done by looking at the semantic content of each and every item that correlates into one factor and then extracting the common core of these items. For instance, if all of the items in a factor ask about hyperactivity/inattention (i.e. “fidgets a lot”, “is easily distracted”) then it is reasonable to assume that these items are assessing one common dimension, since they are conceptually linked and highly correlated. If, however no clear semantic core can be found, or if the factor is littered with a variety of items with little in common, then the items must be reworded or the test re-tested, until clear factors can be found.

In 1994, Robert Goodman conducted a factor analysis on a modified version of the Rutter parent questionnaire and found five factors: Conduct problems, Emotional problems, hyperactivity, peer problems and prosocial behaviour (see Goodman, 1997). In studies examining the properties of these scales the hyperactivity/inattention scale has usually fared best in the parent and teacher editions, whereas the scales measuring peer- and conduct problems have had the least clean factor loadings for teachers and parents (see Agnes Huld Hrafnsdóttir, 2005). These five factors have been the dimensions which the SDQ measures ever since.

Generally speaking there has been good support for the five factor model. In their meta-analysis of the psychometric properties of the parent- and teacher forms of the SDQ, Stone et al (2010) found that 15 of 18 studies supported this model. Goodman (2001) found in his investigation that all 25 items on the parent edition had their highest loadings on the predicted factor and 24 of 25 items on the teacher edition had their highest loadings on the predicted factor. But the factor loadings are not always that clear. In the US, the five factor structure has not proven satisfactory, though support was found for a three-factor structure (Dickey & Blumberg, 2004). This three factor structure contained a prosocial, an externalizing and an internalizing

factor (Dickey & Blumberg, 2004). The proposed three factor solution has not done well in Europe, but the original five factor model has gotten good support (Stone et al, 2010), though Koskelainen, Sourander and Vauras (2001) found a similar three factor structure for the Finnish self-report version.

Becker et al (2006) report that the five factor model of the parent form SDQ explained 47% of the variation when 1,459 children with ADHD were assessed in 10 European countries, including Iceland. There were statistically significant differences between countries, for both total scores and for each subscale. Denmark, France and the United Kingdom generally had the highest scores on difficulties scores whereas Iceland and Switzerland had the lowest (Becker et al,2006).

The Dutch version of the SDQ has also supported the five factor model, where all items had their highest loadings on the predicted factors on the parent version which explained 47,6% of the variation (Muris, Meesters & Van den Berg, 2003). Van den Berg et al (2003) also investigated the Dutch self-report version. It too supported a five factor structure, which explained 43,9% of the variation, but the loadings were not as clean, three (out of five) items which assess conduct problems loaded on other factors and one peer problem item had it's highest loading on the conduct problem factor. The German version of the self-report SDQ also supported a five factor structure than explained 51,4% of the variation (Becker et al, 2004).

The Finnish version of the SDQ self-report form was examined by Koskelainen, Sourander and Vauras in 2001, and they found the factor structure to be satisfactory but not perfectly so. They also did separate investigations on the structure by gender and found that for girls, the loadings for the *emotional problems, prosocial behaviour* and *conduct problems* were in accordance with the original five factor structure, whereas *hyperactivity* and *peer problems* had two and one items respectively that had their highest loadings on other factors. A similar pattern emerged for boys, where (as for girls) *prosocial behaviour* and *emotional problems* had satisfactory loadings and *peer problems* had one wayward highest loading. *Conduct problems*, however, had only one item that had its highest loading on the predicted factor.

When they (Koskelainen, Sourander & Vauras, 2001) examined the factor structure via confirmatory factor analysis, they found a three factor structure, which was quite similar for both boys and girls. The first factor had three hyperactivity items

and two conduct problem items. The second factor had all items that assess prosocial behaviour and was thus a conceptually complete factor. The third and last factor contained all of the emotional problem items and two items that assess peer problems.

The SDQ has been available in Icelandic since 2001 (Obel, et al, 2004) and investigations are ongoing. The factorial structure has been less than perfect in the Icelandic edition. Auður Magnúsdóttir and Berglind Sveinbjörnsdóttir (2004) found in their investigation that the teacher form had two items had their highest loadings on non-predicted factors and the parent form had two erroneous factor loadings as well as three extra factors.

Agnes Huld Hrafnsdóttir (2005) investigated parent and teacher forms for five year old children. For the parent form the items belonging to the hyperactivity and prosocial factors all have appropriate loadings. For the other three factors they had the appropriate loadings but they were not strong enough (less than 0.3) and the factor structure explained 42,5% of the variance.

The teacher form had stronger appropriate loadings where only one item which loaded correctly had a loading under 0.3. However there were four items that had erroneous loadings, two from conduct problems and two from peer problems.

Harpa Hrund Berndsen's (2005) study found that 23 of 25 items loaded correctly in the parent form, though it also gave one additional factor. 22 of 25 items had their highest loadings in the teacher form, though it produced an additional two factors. The six factor parent form structure explained 49,8% of the variance, whereas the seven factor teacher form explained 55,3% of the variance. The explained variance of this structure is quite inconsequential, since without a conceptually important five factor structure the validity of the Icelandic SDQ is in serious jeopardy. The SDQ was constructed to assess these five distinct dimensions, if the Icelandic edition does not, then we cannot use it in any meaningful manner.

Reliability

“In psychological testing (and in measurement generally), a generic term for all aspects of the dependability of a measurement device or test. The essential notion here is *consistency*, the extent to which the measurement device or test yields the same approximate results when utilized repeatedly under similar conditions” - Reber, Allen & Reber (2009), (italics in original).

The way for the SDQ to demonstrate reliability, would be to yield similar results to when the same individual is rated by different raters and at different times, if one can assume that the raters have similar amounts of experience with the child and that the strengths or difficulties are stable over time. The way to test this is by examining the tests interrater reliability and test-retest reliability. Interrater reliability refers to “*The degree to which two or more independent observers agree in their assessment of behaviour”* (Reber, Allen & Reber, 2009). The interrater reliability of the SDQ would be determined by correlating the scores from parents, teachers and self-ratings (if available and applicable).

Test-retest reliability refers to the degree to which test scores are consistent across multiple testings. For example the same test may be administered twice after an appropriate interval (so as to reduce carry-over effects) and then examine the correlation between the two scores.

First, however, the internal consistency must be assessed. Internal consistency is a measure of how various parts of a test are associated empirically (Reber, Allen & Reber, 2009). Practically, it is a measure of how well items in a given part of a test correlate with each other and is generally calculated with Cronbach's alpha (Furr & Bacharach, 2008) This has been used for the SDQ to assess the reliability of individual subscales as well as the reliability of the total difficulties scores.

Internal consistency

Since each scale is derived from a factor analysis (Goodman, 1997) which has been replicated extensively, we can be confident that the scales are fairly homogeneous, i.e. that they assess one general dimension each. However, it is important to assess

the reliability of each scale, that we may know how the ratings have fared. With the SDQ we have five factors and three types of raters, which means that each factor must be assessed with respect to the rater classes.

In a pan-European study of the parent form SDQ in 10 countries (including Iceland), Becker et. al (2006) found Cronbach's α coefficients ranging from α=0,58 to α=0.72. The lowest coefficient was found for the hyperactivity/inattention scale and the highest were found for the prosocial behaviour and peer problem scales. This finding does not appear to be universal, however, since in their investigations of the Hong Kong SDQ, Lai et. al. (2010) found their highest internal consistencies with prosocial behaviour *and* with the hyperactivity/inattention scale. Their lowest internal consistency was found in the peer problems scale. So apart from the prosocial behaviour scale, the highest and lowest internal consistencies were in complete opposites.

It does seem that the parent form SDQ has lower internal consistency than the teacher form. In fact, it seems to consistently produce higher Cronbach's α coefficients. For instance, Stone et. al. (2006) report in their meta-analysis on 26 studies of the SDQ that for the parent form, four of five subscales had a mean α coefficient under 0.7, whereas the teacher form had only one mean coefficient that low (for the peer problems scale). The Dutch SDQ has produced mean α coefficients as high as 0.80 for the teacher form, whereas the same study produced a mean α coefficient of only 0.66 for the parent form (van Widenfelt et.al., 2003).

The SDQ also has a self-report form, but the internal consistencies for this form have not been quite as good as for the informant-rater forms. Becker et. al. (2004) found that for the total difficulties scores on the German self-report SDQ were almost as good as for informant-rater forms (α = 0.78 and α = 0.84 respectively). The self-report form had mostly good internal consistency for each subscale, though the conduct problem scale had an α coefficient under 0.6. Each of the parent- and teacher form subscales had higher internal consistencies than the self-report form. Similar results have been obtained for the Dutch SDQ where the mean α coefficient for all subscales was found to be 0.7 for the parent form and 0.64 for the self-report form (Muris, Meesters & van den Berg, 2003).

The SDQ has also been studied in Iceland with somewhat mixed results. In keeping with other studies, Agnes Huld Hrafnsdóttir (2006) found that the teacher

form had better internal consistency than the parent form (mean α: 0.71 and 0.56 respectively). Though the teacher form only had two subscales with an α coefficient over 0.7 (hyperactivity/inattention & prosocial behaviour). Auður Magnúsdóttir and Berglind Sveinbjörnsdóttir (2004) found α coefficients for total difficulties of 0.88 for the teacher form and 0.79 for the parent form. As has usually been the case, the hyperactivity/inattention subscale had the highest internal consistency for both informant forms, which is hardly surprising since this subscale specifically is firmly rooted in DSM-IV ADHD nosology (Goodman & Scott, 1999). They also found that for teacher forms, the lowest internal consistency was for peer problems, a finding that was replicated one year later by Harpa Hrund Bernsen (2005).

Interrater reliability

Interrater reliability refers to the agreement between raters. That is to say, how much concordance there is in their ratings on a specific dimension or individual. Informant ratings are not ratings of a child's "true" behaviour, but rather assessments based on the raters perceptions of the child or youth given the raters experience with the ratee, so it should not come as a surprise that these ratings are not always highly correlated. In fact, in Achenbach's, McConaughy and Howell's (1987) meta-analysis has shown that interrater reliabilities are often around 0.20-0.27 and about 0.60 for the same class of informant (two teachers or two parents).

But these potentially low correlations are not necessarily bad, since they provide information about when or where problematic behaviours occur. If a child displays behavioural difficulties in school but not at home, then we know that there might be a situational factor in the school setting that elicits this difficulty. If perfect correlations were found between rater scores, then we have some assurance that the ratings might be correct. But what we gain in assurance, we lose in diversity. If we have two identical cross-rater ratings then the second one does not add much to our knowledge of the difficulties of the child or youth, it increases reliability, but it may decrease construct validity (Smith, 2007). For example; if a teacher and parent rate the child in an identical fashion on hyperactivity we have perfect reliability on that subscale for this particular individual. However, if self-ratings from the child are available and he/she report less disrupting behaviours and more symptoms of

emotional distress, then we have less reliability but more construct validity (Smith, 2007)

As each type of rater has different experiences with the child or youth, it is unsurprising that each type of rater also tends to differ in accuracy of predictions of specific types of problems. For example; informant ratings tend to give better predictions when it comes to externalizing disorders than do self-ratings, especially in clinical or criminal samples. However, informants tend to underestimate the severity of internalizing disorders, so it's probably best to give the most weight to self-ratings in those circumstances, since their ratings tend to give the best predictions for those types of disorders (Goodman et. al., 2000; Smith, 2007).

As said earlier, the Achenbach, McConaughy and Howell (1987) meta-analysis found a mean interrater reliability of 0.27. The SDQ, however has fared better than that. The NCTSN (www.nctsn.org) report an overall interrater reliability of 0,37 with a maximum of 0.48 for teacher-parent correlation on the hyperactivity/inattention scale. This scale seems to have the highest interrater reliability. Van Widenfelt et. al. (2003) found that the hyperactivity/inattention scale had the highest interrater reliability for all combinations of informants.

The study by Agnes Huld Hrafnsdóttir (2006) also found that the hyperactivity/inattention scale had the highest parent-teacher interrater reliability, though it was only at 0.38, and that the lowest was for prosocial behaviour (0.17)

Stone et. al. (2010) conducted a meta-analysis on all of the SDQ subscales and found correlations of 0.26-0.47 for self-reports and parent raters. The hyperactivity/inattention scale had the highest mean interrater reliability of 0.47 and all subscales except prosocial behaviour had a mean interrater reliability above 0.27. The low interrater reliability of the prosocial scale has also been found by Muris et. al. (2003) and van Widenfelt et.al (2003).

Test-retest reliability:

Symptoms of psychological distress can be more or less temporary, and then there is the issue of spontaneous recovery. This means that symptoms which are reported at one time may lose severity or disappear completely after some time has passed. This is especially true when assessing children, since they sometimes simply "grow out of

it", though this seems less likely for the more severe cases. Unlike alternate forms reliability, in which different parts of a test are treated as separate tests and then correlated, one can be confident that the content does not differ when administering the same test twice. However, we must assume that the traits/symptoms we assess remain stable over time (Furr & Bacharach, 2008).

The SDQ has fared well in investigations of its test-retest reliability, and the results indicate that the teacher form has the best temporal consistency. Goodman (2001, see Agnes Huld Hrafnsdóttir, 2005) reports a test-retest reliability of 0,72 for parents and 0.80 for teachers in a 6 month interval. The superior test-retest reliability of the teacher form was confirmed in Stone et. als (2010) meta-analysis of 6 studies, where the teacher form had better reliability on all subscales, the total difficulties score and on the impact score. The overall mean test-retest reliability was 0.67 for parents (only passing 0.7 on the hyperactivity/inattention scale and total difficulties score), whereas the overall mean for teachers was 0,77, dropping below 0.7 on the impact score only.

Validity

"In testing, of any measuring instrument, device or test, the property of measuring that which it is purported is being measured" - Reber, Allen & Reber (2009). In order for the SDQ to be a valid screening instrument it must accurately identify children and youths who are having difficulties. In order to determine its validity, some external criteria must be enforced, that is, it must be compared to other, established measures.

Predictive Validity

"Predictive validity evidence refers to the degree to which test scores are correlated with relevant variables that are measured at a future time" (Furr & Bacharach: 2008). Predictive validity is a form of convergent validity, since its strength is dependent on the match between test scores and other relevant variables. For the SDQ, this means that the strength of its predictive validity is dependent on the extent to which scores will correlate with the occurrence of actual psychopathology**.**

The SDQ has an algorithm for identifying mental disorders. It does so by combining information from multiple informants with both symptom- and impact scores. The algorithm will identify disorders of three general kinds: Conduct-oppositional disorders, hyperactivity-inattention disorders and anxiety-depressive disorders. Each kind is then categorized as being either unlikely, possible or probable (Goodman et al. 2000).

An exceptionally important aspect of identification is correctness, for what's the point of making incorrect identifications? The correctness of an identification is assessed by sensitivity and specificity. Sensitivity is a measure of a tests ability to give a correct positive.(Furr & Bacharach, 2008). In the case of the SDQ, a high sensitivity means that it is good at correctly indicating that a child or youth is at risk of having a disorder. Sensitivity can also be expressed as the the ratio of True Positives to False Negatives, that is the ratio of correctly given positives to incorrectly given negatives. A True Positive is when someone with a disorder is categorized as having that disorder by the test. A False Negative is when someone with the disorder is

incorrectly deemed not to have it (Furr & Bacharach, 2008).Specificity is the measure of a tests ability to give a correct negative For the SDQ this means that high specificity would indicate that it will correctly categorize evaluees as not having a given disorder, by for example putting individuals in the unlikely group instead of the likely or probable groups. Specificity can be expressed as how the test categorizes the evaluees who in reality do not have the disorder. These can be False Positives, in which they are deemed to have a disorder by the test, or they can be True Negatives, in which they are correctly deemed not to have a disorder (Furr & Bacharach, 2008).

A test's Positive Predictive Value has to do with the tests ability to correctly give positive scores in a given sample. It is the proportion of True positives to the number of positive calls (True positives + False positives). A perfect Positive predictive value is one in which all positive calls are True positives (Furr & Bacharach, 2008). A tests Negative Predictive Value is the proportion of True negatives to the number of negative calls (True negatives+False negatives). A perfect Negative Predictive Value is one in which all negative calls are True Negatives (Furr & Bacharach, 2008).

Goodman et al. (2000) conducted a rigorous investigation on the SDQs algorithm's ability to give correct positives and negatives by comparing it's results with a community sample who had also been reviewed by independent interviewers. Based on this information the subjects were given ICD-10 diagnoses by clinical raters when appropriate. They found the sensitivity to be 63,3%, specificity to be 94,6%, positive predictive value to be 52,7% and negative predictive value to be 96,4%. They also examined the algorithms sensitivity to detect specific disorders with varied results. For conduct, depressive, hyperactivity and some anxiety disorder the sensitivity was over 70% with hyperkinetic disorder peaking at 86,1%. Other disorders were not as readily discovered though, bottoming out with specific phobias at 30,9% (Goodman et. al, 2000)

In the same study, Goodman et. al. (2000) found that teacher evaluations were better than parent evaluations in detecting externalizing disorders, though this difference was only significant for conduct disorder. Parent evaluations were better at predicting internalizing disorders, though this difference was only significant for anxiety disorders. Self-ratings were found to be the least sensitive measures, though

they proved to be more useful than parent- or teacher ratings at detecting some emotional disorders. Combining all of the ratings gave the best result. Sensitivity for any psychiatric disorder for Parent+Teacher+Self-report was 64.8%, Parent+Teacher was 59.4%, Teacher+Self-report was 47.9%, Parent+Self-report was 41.3%, , Teacher-ratings was 38.7%, Parent-ratings was 33,7% and Self-report was 15.9% (Goodman et al, 2000).

Becker at al (2004) however, found that self-reports were as good as parent reports at overall prediction on the German version of the SDQ. Although both teacher and parent ratings were better at predicting specific subgroups than self report measures. They also found that generally speaking, the adult ratings together gave a better prediction than self-reports with an adult rating, a finding similar to that of Goodman et al (2000). Becker at al (2004) also concluded that if ratings from both parents and teachers were available, then self-ratings did not significantly add to the predictive power of the list. So based on this result, and of the recommendation of Goodman et al (2000), if you had to do without one informant rating, you're better off skipping self-reports than adult ratings.

Congruent validity

"A method of establishing the validity of a new test by correlating scores from it with scores from another test of established validity" - Reber, Allen & Reber (2009).

The Rutter questionnaires and the Child Behavior Checklist (CBCL) have been extensively used in the realm of child evaluation. These have been the standard against which the SDQ has been and must be measured. Overall the SDQ has done well in the investigations of its congruent validity against these two established measures. If the SDQ is as good a measure as the Rutter questionnaires and the CBCL, then it would be a preferable measurement since it is much shorter than the CBCL (SDQ 25 items, CBCL 118 items) and should be better received by raters since it asks about strengths as well as difficulties (Goodman & Scott, 1999).

The SDQ and Rutter Questionnaires

Goodman developed the SDQ from a modified version of a parent form Rutter questionnaire, so it is interesting to see that in Robert Goodman's (1997) comparison of the lists abilities to discriminate between a clinical and non-clinical sample was exactly the same for the parent versions (area under the curve of 0.87 for both lists) and for the teacher versions (0.85 for the SDQ and 0.84 for the Rutter questionnaires). This was established using a statistical technique called Receiver Operating Characteristics (ROC) curve (Goodman, 1997). An area under the curve of 1 indicates perfect discrimination, an area under the curve of 0,5 indicates a discrimination no better than chance. In the same study Goodman (1997) also examined the correlations between scores of the two lists parent and teacher versions on comparable dimensions. The Rutter questionnaires do not assess peer problems or prosocial behaviour and these dimensions could thus not be compared. The remaining factors (total score, conduct problems, emotional symptoms and hyperactivity) were compared and the correlations were convincing, ranging from 0.92 (teachers total scores) to 0.78 (parents emotional symptoms) with teacher versions correlating higher on all dimensions (Goodman, 1997).

The SDQ and the Child Behavior Checklist

The correlations between the Child Behavior Checklist (CBCL) and the SDQ have shown these measures to correlate sufficiently. Correlations of total difficulties scores have ranged from r = 0.70-0.87 for parent-rating and r = 0.68-0.87 for teacher-ratings (Stone et al, 2010). Subscale correlations on comparable factors have been sufficient ($r > 0.7$) for externalizing and hyperactivity, attention problems and conduct problems, but less-than-sufficient ($r < 0.7$) for emotional problems, peer problems, social problems and internalizing problems (Stone et al, 2010). No comparison could be made for prosocial behaviour since the CBCL does not inquire about such behaviour.

Goodman and Scott (1999) found total score and subscale correlations on parent forms from r = 0.59 for social/peer scales to r = 0.87 for total score. Van Widenfelt et al (2003) found that the Dutch version of the SDQ parent form and CBCL Dutch parent form had correlations ranging from r = 0.48 (Peer problems/withdrawn) to r = 0.78 (hyperactivity/attention problems) on comparable

subscales. The correlation of parent forms total scores on SDQ and CBCL was r = 0.74. Goodman and Scott's (1999) investigation of the SDQ and CBCL showed that both measures were equally good at discriminating scores from a clinical and a non-clinical sample, with an area under the curve (AUC) of 0.93 for the SDQ and 0.92 for the CBCL on total scores. The two measures differ significantly in their ability to discriminate on any comparable subscale.

So the SDQ is not only as good at rating children and youths as the Rutter questionnaires and the CBCL, but it is also shorter and more pleasant to fill out. In fact, in Goodman and Scott's (1999) study on the comparability of the SDQ to the CBCL, they found that 41 out of 64 parents who expressed a preference preferred the SDQ to the CBCL. This difference was statistically significant, though only parents from a non-clinical sample were asked which measure they preferred. This makes it an obvious choice as a first, overall screening measure. Goodman and Scott (1999) suggest that because of its brevity and because it seems to be better received by low-risk sample parents at least, it could prove a useful community screening instrument. By being shorter and not solely assessing negative features, it may have better response rates than if screening by longer and more negatively worded measurements. They also suggest that because of its more general nature, the SDQ would be more valuable as an initial screening instrument, later succeeded by the CBCL in further investigations, since it covers a wider selection of problems and rarer disorders than the SDQ.

Aims of the study

This study is part of a greater research project which aims to hone the psychometric properties of the Icelandic translation of the Strengths and difficulties questionnaire. Previous psychometric inquiries into the properties of the Icelandic SDQ have been unsatisfactory. The factor analyses have produced extraneous factors, factor loadings have been in violation of the original five-factor structure and reliabilities have been sub-par. The new translation has changed the wordings of five items (described in detail in the Method section), with the hope of producing results more in keeping with the lavishly reproduced factor structure and loadings provided by Robert Goodman.

Method

Participants

Participants were 16 teachers from 3 schools. Head masters from 10 schools were contacted and 4 Head masters agreed to have their schools participate. School response rate = 40%. No teachers from Landakotsskóli agreed to participate.

No special consideration was given to stratify the sample based on Socio-economic-status (SES), rural vs. urban environment or any other variables. Schools were simply contacted once they had been identified as suitable for the study (teaching 6-10 year-olds).

Each teacher who taught 6-10 year-olds from schools whose head master had agreed to let the teachers participate was contacted and asked for participation. Each teacher was eligible to take part in a lottery to win a 25 000 ISK gift certificate for participating. This was thus a convenience sample. The teachers rated 80 children in all, 42 boys and 38 girls. The mean age for girls was 8,37 and 8,32 for boys. The mean age was 8,34 in all.

Measurement

The current Icelandic translation of the SDQ teachers form along with the Icelandic translation of the impact supplement was used and tested (see appendix A). The previous translation (available on http://sdqinfo.org) was found to be unsatisfactory in earlier studies and five items were changed:

Old	New
Deilir greiðilega með öðrum börnum (nammi, dóti, blýöntum o.s.frv.	Á auðvelt með að deila með öðrum börnum (nammi, dóti, blýöntum o.s.frv.
Fær oft skapofsaköst eða er heitt í hamsi	Fær oft skapofsaköst eða er uppstökkur
Almennt vel þokkaður/þokkuð af öðrum börnum	Öðrum börnum líkar almennt vel við hann/hana
Auðvelt að stela athygli hans/hennar, einbeiting á flakki	Truflast auðveldlega, hann/hún á erfitt með að einbeita sér
Óörugg/ur, hangir í foreldrunum við ókunnar aðstæður, missir sjálfstraust	Óörugg/ur hangir í foreldrunum við ókunnar aðstæður, missir auðveldlega sjálfstraust

These changes may seem minor, but as Goodman and Scott (1999) have shown, a small difference in wording can actually have a large effect on the results. These differences can hardly be smaller when translating from one language to another, so it's important to word the items carefully. It does not seem likely that the unsatisfactory results from the investigations of the Icelandic SDQ be due to some fundamental cultural differences, since results from other Nordic countries have been quite satisfactory (Obel et al, 2004) and these cultures are very similar.

No changes were made to the impact supplement.

Procedure

16 teachers were given the current Icelandic translation of the SDQ teachers form and impact supplement. They were asked to assign numbers to the students in their class and provide the experimenter with these numbers. The experimenter then chose 5 numbers for each class at random using a random number generator found at www.random.org, thus choosing who was to be rated. The experimenter had no access to any personal information since no names were used, only numbers. The parents of the students were sent an email written by the experimenters instructors (see appendix C) in which they could give their informed consent to their child's participation. If consent was given, the teacher proceeded to rate the 5 children accordingly and sent the lists back to the experimenter. If parents refused to give their consent, then another student was to be chosen using the same method. Although no parents/guardians refused participation.

Statistical analysis

All statistical analysis was conducted in SPSS 19 for Windows. Means and standard deviations for each subscale and the total difficulties score were calculated. Internal consistencies were assessed with Cronbach's alphas (α) and inter-item, item-total, item-total-difficulties and inter-scale correlations are reported and discussed below. Finally an exploratory factor analysis (EFA) was conducted using Principal component analysis with a Varimax rotation like Robert Goodman did in his study from the year 2000 (see Agnes Huld Hrafnsdóttir, 2005). Another exploratory factor analysis was conducted after the scree plot indicated a five factor structure instead of the eight that the first EFA produced. This second EFA was predetermined to draw

five factors, since the scree plot suggested it and the questionnaire is designed to assess five dimensions (Goodman, 1997).

Results

Descriptive statistics

Means and standard deviations were calculated for each of the subscales and the total difficulties scores. Despite the small sample size, the data are quite close to the UK norms. These are presented in table 2 below.

Table 2 - Means and standard deviations of the current study and UK norms

	Teacher SDQ (N= 80)	Teacher SDQ (N = 8208)
	Icelandic means (standard deviations)	**UK means (standard deviations)**
Total	6.3 (5.5)	6.6 (6.0)
Emotional	1.2 (1.5)	1.4 (1.9)
Conduct	1.1 (1.6)	0.9 (1.6)
Hyperactivity	3.0 (2.9)	2.9 (2.8)
Peer problems	0.9 (1.6)	1.4 (1.8)
Prosocial	7.8 (2.4)	7.2 (2.4)

Scores on the total difficulties scores ranged from 0-21 out of 0-40 possible with only one subject scoring 21. 53.2% scored 5 or lower and 16.5% scored 0. Goodman (1997) suggested that a score of 17-40 on this scale should indicate abnormality, and that these scores should capture only the top 10% of scores. The top 94.9% in this study scored 18 (no one scored 17 or even 16). This only captures the top 5.1% of the sample. If the 90th highest percentile is used instead, then a score of 14 indicates abnormality (top 91.1 percentile). However, in Goodman's (1997) recommendations a score of 14 makes up the very bottom score on the borderline range.

The emotional symptoms subscale had a mean of 1.2 and a standard deviation of 1.5. The means for the individual items ranged from 0.18 (*Óttast margt, verður auðveldlega hrædd/ur; Oft óhamingjusamur/söm, langt niðri eða tárast*) to 0.33 (*Áhyggjur af mörgu, virðist oft áhyggjufull/ur*). 63.8% of endorsements were on

the interval 0-1 and the other 46.2% between 2-5. Only 3 individuals received the highest score of 5 (out of a possible 10), whilst 37 children (of 80) received a score of 0. The use of the 90th percentile to indicate abnormality in this scale landed on a score of 3, which would include 17 individuals. The suggested bandings provided by Goodman (1997) puts a score of 3 in the Normal range, whereas a score of 6-10 should indicate abnormality.

The conduct problems subscale had a mean of 1.1 and a standard deviation of 1.6. The means of the individual items ranged from 0.03 (*Stelur heima, í skóla eða annars staðar*) to 0.43 (*Almennt hlýðin/n, gerir yfirleitt eins og fullorðnir óska*). A score of 0 was endorsed 57% of the time and scores of 1 and 2 were endorsed 16.5% and 10.1% of the time respectively. Only 1 individual received the highest score of 7 (out of 10 possible), which was also true for a score of 6. A score of 3 corresponds to the 89.9th percentile in this sample on this scale. In Goodman's (1997) bandings a score of 3 should be in the borderline range.

The hyperactivity scale had a mean of 3 and a standard deviation of 2.9. The means for each individual item ranged from 0.44 (*Stöðugt með fíkt eða á iði*) to 0.76 (*Fylgir verkefnum eftir til enda, heldur góða athygli*). Scores between 0 - 3 were endorsed 61.3% of the time with an endorsement range of 0-10 (out of 10 possible). The most frequently given score in this scale was 0, given to 23 individuals. Goodman's (1997) suggested bandings say that a score of 7-10 would indicate an abnormal range on this scale, encompassing the top 90th percentile. A score of 7 in this sample was obtained by the 91.3rd percentile, with eight individuals receiving a 7, three receiving an 8, two receiving a 9 and two receiving 10's. So this scale at least is very much in keeping with Goodman's normative data and bandings.

The Peer problems scale had a mean of 0.9 and a standard deviation of 1.6. The means of the individual items ranged from 0.11 (*Semur betur við fullorðna en önnur börn*) to 0.36 (*Öðrum börnum líkar almennt vel við hann/hana*). The scores 0 and 1 together were endorsed 73.8% of the time with the score 0 being the most frequently endorsed or 61.3% of the time. A score of 3 or above was obtained by the top 92.5% of scorers in this scale which corresponds to the normal range, instead of the abnormal range, as proposed by Goodman in 1997. The highest score given was 7, which together with the score of 6, were the least endorsed scores given to only one individual each.

The prosocial scale had a mean of 7.8 and a standard deviation of 2.4. The individual subscale item means ranged from 1.3 (*Býðst oft til að hjálpa öðrum (foreldrum, kennurum, öðrum börnum*) to 1.8 (*Góð/ur við yngri börn*). The most frequent score given in this scale was 10, which is the highest score. It was given to 21 subjects or 26.3% of the sample data. Goodmans (1997) bandings suggest that a score from 0-4 should indicate prosocial abnormality and a score of 5 should indicate borderline difficulties. In this sample, the bottom 8.8% scored a 4 or less and 15% scored 5 or less, which seem to be a fair approximation of Goodmans guidelines.

Factor analysis

Factor analyses were conducted to examine the factor structure of the Icelandic SDQ. Firstly an exploratory factor analysis was conducted with no limitations on how many factors were to be drawn, even though there is enough theoretical basis to justify a factor analysis that draws only 5 factors. When Goodman himself was examining the factor structure of the SDQ he used an exploratory factor analysis with principal component analysis and varimax rotation (see Agnes Huld Hrafnsdóttir, 2005). Thus the factor analysis conducted here will follow the same format. The factor analysis produced 8 factors with Eigenvalues over 1 which explained 71.85% of the variance. The first factor explained 24.03% of the variance, the second 10.64%, the third 9.21%, the forth 7.79%, the fifth 6.65%, the sixth 4.98%, the seventh 4.51% and finally the eighth 4.02% of the variance. The factor correlations of the factor analysis, along with the explained variance of each factor are presented in table 3 on the next page

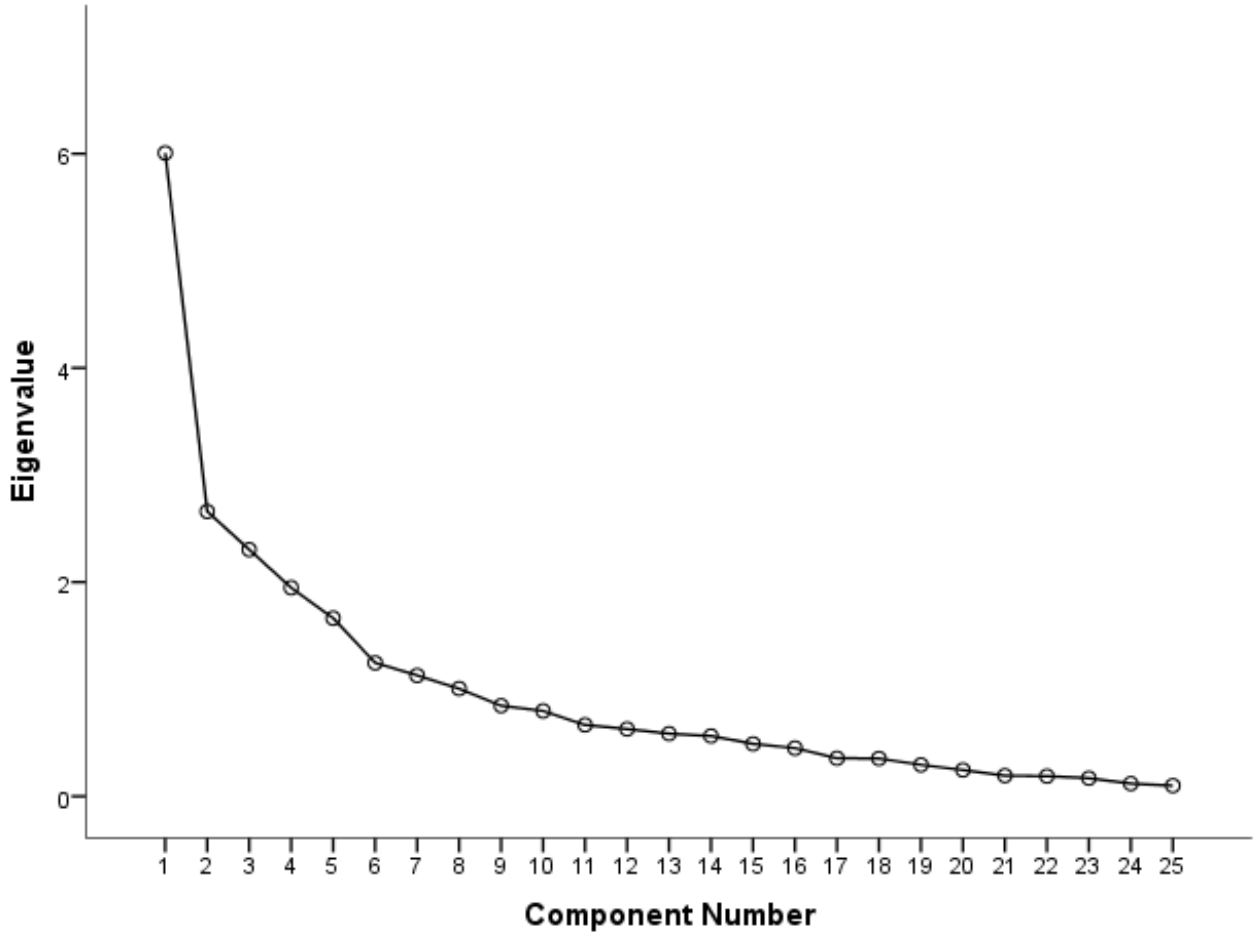
As can be seen in the table, the hyperactivity- and peer problems scales (Factors 1 & 2) were both perfectly reproduced with satisfactory correlations. Factor 3 contained 3 emotional symptoms items and 1 conduct problem item (*Skapofsaköst*), totalling in four items. Factor 4 consisted of three items which were all from the prosocial scale with factor correlations around 0.8. Factor 5 also consisted of only 3 items, though these were all from the conduct problems scale. These were; *Flýgst of á eða leggur börn í einelti*; *Almennt hlýðinn gerir eins og fullorðnir óska* & *Lýgur oft eða svindlar.* Factor 6 had only 1 high positive loading (*Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* -an emotional symptoms item)

of 0.825. It also had some minor secondary loadings and negative loadings, but nothing conclusive or interesting. Factor 7 consisted of three items, two negative loadings from the prosocial scale (*Á auðvelt með að deilia með öðrum börnum.. & Tekur tillit til tilfinningar annarra*) as well as a positive loading from the conduct problems scale (*Stelur heima, í skóla eða annars staðar*). Since the prosocial items are correlated negatively with this factor and the stealing-item has a positive correlation, it seems to deal with inconsiderate property-behaviour. A child who scores highly on this particular factor would probably be considered greedy and inconsiderate. The 8:th and final factor contained only the item *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika.*

The scree plot (graph 1) from this factor analysis is presented on the page after the table.

Table 3. Results of factor analysis with principal component analysis and varimax rotation

	Hyperactivity	Peer Problems	Emotional & temper tantrums	Prosocial	Conduct problems 1	Insecure	Negative prosocial and steals	Complains.
Truflast auðveldlega	**,841**				,185			,149
Fikt og Íði	**,825**		,237		-,104	-,142		
Eirðalaus	**,812**	,155			,210			
Fylgir verkefnum	**,782**	,173		-,234		,125		
Hugsar áður en framk.	**,567**			-,295	,267		,327	
A.m.k einn góðan vin		**,809**		-,182				
Semur betur fullorðna	,159	**,679**	,187			,152		
Verður fyrir stríðni	,144	**,672**			,265	-,311		-,145
Frekar einræn/n		**,638**		-,144		,444		,240
Öðrum börnum líkar	,190	**,622**		-,131	,200		,351	-,224
Óttast margt,			**,828**		,150	,180		
Óhamingjusamur/söm		,212	**,788**				,330	
Áhyggjur af mörgu	-,167	,252	**,622**	,208	,129	,367	-,124	,245
Fær oft skapofsaköst	,268	,271	**,579**		,321	-,457		-,106
Hjálpsamur	-,198			**,809**	-,145		-,166	,123
Góður við yngri börn		-,227	-,115	**,804**				,100
Býðst til hjálpa				**,801**			-,203	-,134
Flýgst á og einelti	,143	,127			**,804**			
Almennt hlýðinn	,310	,127	,104	-,177	**,663**		,323	
Lýgur eða svindlar			,547	-,333	**,587**	-,240		,113
Óöruggur í ókun. aðs.	,127	,108	,134			**,825**		-,193
Á auðvelt með deila			-,160	,313		-,160	**-,658**	
Tekur tillit		-,185		,164	-,300		**-,589**	
Stelur heima í skóla...		,204	,283	,357	-,304	-,288	**,429**	-,240
Kvartar um höfuðverk.	,186					-,116		**,910**
Explained variance	**24.03%**	**10.64%**	**9.21%**	**7.79%**	**6.65%**	**4.98%**	**4.51%**	**4.02%**



Graph 1. Scree plot from exploratory factor analysis 1

Even though the factor analysis produced 8 factors with Eigenvalues over 1, the scree plot (Graph 1) it produced seemed to suggest a five factor structure. The scree plot isn't exceptionally clear, it could almost as easily be eight factors as it is five. However, since there is a good theoretical basis for a five factor structure (Goodman, 1997), another factor analysis was conducted in which five factors were to be drawn. The decision to draw five factors is in accordance with the theoretical basis behind this questionnaire. Even though the first factor analysis gave 8 factors, it could still prove useful to limit the amount of possible factors to the predetermined 5, since the questionnaire is designed to measure 5 dimensions.

Once again a principal component analysis with varimax rotation was used. The five factor solution explained 58.32% of the variance, with each factor explaining

24.03%, 10.64%, 9.21%, 7.79% and 6.65% of the variance respectively. The results from the 5-factor analysis are presented in table 4 below.

Table 4. Results of factor analysis with varimax rotation with 5 factors determined to be drawn.

Item	Hyperactivity + obedient + complains	Prosocial	Emotional & conduct	Peer problems and Steals	Insecure & solitary
Truflast auðveldlega	**,872**				,114
Eriðarlaus	**,832**			,159	
Fylgir verkefnum	**,767**	-,179	-,117	,127	,154
Fikt og iði	**,734**		,133	,254	-,161
Hugsar áður.. framkvæmir	**,598**	-,404	,185		
Almennt hlýðinn	**,444**	-,402	,379		,153
Kvartar oft um höfuðverk..	**,313**		,165	-,295	
Hjálpsamur	-,181	**,811**			
Býðst til að hjálpa		**,802**			
Góður við yngri börn		**,767**		-,197	
Á auðvelt með að deila	,128	**,486**	-,238		
Tekur tillit	-,129	**,405**	-,265	-,214	
Óttast margt			**,773**		,139
Oft óhamingjusamur			**,727**	,327	
Lýgur eða svinlar	,201	-,382	**,713**		
Fær oft skapofsaköst	,292		**,649**	,397	-,300
Áhyggjur af mörgu	-,134	,239	**,602**		,512
Flýgst á eða leggur í einelti	,338	-,196	**,410**		,215
Á a.m.k einn góðan vin		-,215		**,746**	,271
Verður fyrir stríðni	,201	-,125	,174	**,670**	
Öðrum börnum líkar við.	,197	-,303	,104	**,666**	,123
Semur betur við fullorðna	,186		,189	**,546**	,399
Stelur heima, í skóla...		,278	,209	**,519**	-,320
Óöruggur við ókunnar aðst					**,750**
Frekar einræn/n		-,172		,379	**,659**
Variance explained	**24.03%**	**10,64%**	**9,21%**	**7,79%**	**6,65%**

Even though the previous scree plot had indicated a five factor structure, the factor analysis did not replicate the five factor structure proposed by Goodman (1997).The first factor contained all hyperactivity items with correlations ranging from 0.598-0.872. It also contained 1 conduct problem item (*Almennt hlýðinn, gerir eins og fullorðnir óska*) and 1 emotional problem (*Kvartar oft um höfuðverk, magaverk eða*

flökturleika oft um höfuðverk, magaverk eða flökturleika), with the correlations 0.44 and 0.31 respectively. Factor 2 consisted of all of the prosocial items with correlations ranging from 0.40-0.81, thus the prosocial scale was reproduced. Factor 3 was a six-item factor that consisted of 3 emotional items and 3 conduct problem items. The emotional problems were *Óttast margt, verður auðveldlega hrædd/ur; Oft óhamingjusamur/söm, langt niðri eða tárast* and *Áhyggjur af mörgu, virðist oft áhyggjufull/ur* with correlations ranging from 0.60-0.77. The conduct problem items were *Lýgur oft eða svindlar; Fær oft skapofsaköst eða er uppstökkur* and *Flýgst oft á eða leggur börn í einelti* with correlations ranging from 0.41-0.71. Factor 4 consisted of 4 peer problem items with correlations ranging from 0.54-0.74 and 1 conduct problem item (*Stelur oft heima, í skóla eða annars staðar*) with a correlation of 0.52. The fifth and final factor consisted of only two items, one emotional problem (*Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust*) and one peer problem (*Frekar einræn/n leikur sér oft ein/n*) with the correlations 0.75 and 0.66 respectively.

Internal consistencies and inter-item, item-total correlations

In the table below (Table 5) are the inter-scale and total difficulties correlations along with the Cronbach's alpha for each subscale and the total difficulties score. The duplicate correlations have been removed for the sake of simplicity.

Table 5. Inter-scale and total difficulties correlations and internal consistencies

	Emotional symptoms	Conduct problems	Hyperactivity	Peer problems	Prosocial	Total difficulties
Emotional symptoms	1					
Conduct problems	0,398	1				
Hyperactivity	0,165	0,422	1			
Peer problems	0,226	0,452	0,363	1		
Prosocial	-0,09	-0,439	-0,314	-0,366	1	
Total difficulties	0,541	0,760	0,811	0,674	-0,431	1
α	0,52	0,71	0,86	0,76	0,69	0,84

The inter scale correlations for difficulties range from 0.165 to 0.452 and all correlations with the prosocial scale are negative. Total difficulties correlations range from -0.431 to 0.811.The corrected item-total correlations and Cronbach's Alpha if item deleted are presented below in table 6.

Table 6. Corrected item-total correlations and Cronbach's Alpha if item deleted for each subscale.

	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Emotional Symptoms		
Kvartar	-,007	,653
Áhyggjur	,525	,275
Óhamingjusamur	,439	,385
Óöruggur	,168	,538
Óttast margt	,467	,370
Conduct problems		
Skapofsaköst	,544	,635
Hlýðinn	,579	,629
Flýgst á eða leggur í einelti	,516	,647
Lýgur	,646	,596
Stelur	,046	,763
Hyperactivity		
Eirðalaus	,741	,812
Fikt eða iði	,666	,832
Truflast auðveldlega	,760	,807
Hugsar áður en framkvæmir	,536	,862
Fylgir verkefnum til enda	,683	,828
Peer problems		
Einrænn	,463	,745
A.m.k. einn góðan vin	,655	,680
Öðrum líkar við	,566	,717
Verður fyrir stríðni	,542	,716
Semur	,511	,734
Prosocial behaviour		
Tekur tillit til tilfinninga annarra	,353	,740
Á auðvelt með að deila með öðrum	,361	,684
Hjálpsamur/söm...	,644	,578
Góð/ur við yngri börn	,564	,639
Býðst oft til að hjálpa öðrum	,549	,607

The emotional symptoms scale had a Cronbach's alpha of 0.52 and item-total correlations ranging from -0.007 (*Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika*) to 0.525 (*Áhyggjur af mörgu, virðist oft áhyggjufull/ur*) with a mean item-total correlation of 0.318. This scale had some problematic items, the most notable of which was *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika.* This item invariably had low or negative correlations to other items in the subscale. Its lowest correlation of -.148 was with *Óöruggur* and its highest was with *Áhyggjur* (0.11).

The low correlations for *Kvartar oft um höfuðverk, magaverk eða flökturleika* was probably not due to low endorsement rates caused by the small sample size. 15% of endorsements (12 instances) received a rating of 1, and 6,3% of endorsements were for a score of 2. So some children were rated as moderate to somewhat persistent complainers. The item *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* was one of the items that had been changed in this translation. It had 4 correlations under 0.3, though one of them was with the problematic item *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika*. *Óörugg/ur* did have an unsatisfactory correlation with the subscale total of 0.168, but a correlation of 0.27 with total difficulties.

The conduct problems scale had a Cronbach's alpha of 0.71 and item-total correlations ranging from 0.046 (*Stelur heima, í skóla eða annars staðar*) to 0.646 (*Lýgur oft eða svindlar*), with a mean item-total correlation of 0,468. All items except for *Stelur heima, í skóla eða annars staðar* had an item-total correlation over 0.5. In fact, the Cronbach's alpha would increase to 0.763, were this item deleted. It was the only item in the scale that would increase the subscales alpha, if it were deleted. This item had the lowest correlations in this scale. Its highest inter-item correlation was 0.25 and its lowest was -0.07. Its correlation with total difficulties was 0.12. The item that had been changed from previous translations (*Fær oft skapofsaköst eða er uppstökkur*) had a mean inter-item correlation of 0,38 and an item-total correlation of 0.544. It also had a correlation of 0.6 with total difficulties.

The hyperactivity scale had a Cronbach's alpha of 0.86 and thus had the highest internal reliability. Its item-total correlations ranged from 0.536 (*Hugsar áður*

en hann/hún framkvæmir) to 0.760 (T*ruflast auðveldlega, hann/hún á erfitt með að einbeita sér*), with a mean item-total correlation of 0.677. The inter-item correlations ranged from 0.36 to 0.69 with only two correlations under 0.4. This factor was perfectly reproduced in the 8-factor analysis. In the 5-factor analysis all of the hyperactivity items made up one factor along with two stray items from other subscales. The item that had been changed in this translation *Truflast auðveldlega...* had the highest item-total correlation and among the highest inter-item correlations within the subscale. It also had an item-total- difficulties correlation of 0.86.

The peer problems scale had a Cronbach's alpha of 0.76 and item-total correlations ranging from 0.463-0.655 with a mean item-total correlation of 0.547. The inter-item correlations ranged from 0.30 to 0.50. This subscale was perfectly reproduced in the 8-factor analysis but not in the 5-factor analysis. In the 5-factor analysis the item *Frekar einræn/n, leikur sér oft ein/n* had only one correlation over 0.1, which was with an emotional problems item. These two items made up one factor. The other four items in the subscale made up one factor together with a conduct problems item (*Stelur oft heima, í skóla eða annars staðar*) in the 5-factor analysis. The inability of the item *Einrænn* to correlate with a peer problems factor in the 5-factor analysis is reflected in the fact that it had only one inter-item correlation over 0.4 within the scale, and its other three correlations in the area of 0.3. The item that had been changed from previous translations, *Öðrum börnum líkar almennt vel við hann/hana* had an item-total correlation of 0.566 and a mean inter-item correlation of 0.42. It also had a correlation of 0.55 with the total difficulties score.

The prosocial subscale had a Cronbach's alpha of 0.69 and item-total correlations ranging from 0.353-0.644 with a mean item-total correlation of 0.494. The inter-item correlations ranged from 0.23-0.65. The item *Á auðvelt með að deila með öðrum börnum...* was the last of the five items that had been altered in this translation. It had an item-total correlation of 0.361 and inter-item correlations ranging from 0.23-0.34 with a mean inter-item correlation of 0.27. Its correlation with the total difficulties score was -0.17.

Discussion

Descriptive statistics

The results section presented the 90:th percentile scores to indicate abnormality for this sample and compared it to Goodman's normative data presented in his 1997 study (Goodman, 1997). However, little can be drawn from a comparison of these data, since Goodman's (1997) normative data comes from a sample of 4-16 year-olds, not just the 6-10 year-olds in this sample. Also the sample in this study consisted of only 80 ratings, as opposed to the 403 in Goodman's (1997) study. The comparisons were still deemed to be necessary, since they may prove useful for further studies.

On the emotional symptoms subscale, the 90:th percentile landed on a score of 3, which in Goodman's (1997) data would be in the Normal range. This effect is most likely due to the small sample size more than anything else, though as mentioned in the results section, the item *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika* didn't correlate well with its subscale. Which may indicate that is was not considered a symptom of the same kind of difficulties as the other items in the emotional symptoms subscale by some of the teachers.

On the conduct problems subscale a score of 3 was obtained by the top 89,9% of scorers. Which would be in the borderline range in Goodman's (1997) suggestions. This scale faced similar problems as the emotional symptoms subscale in that one item (*Stelur heima, í skóla eða annars staðar*) didn't correlate well with the other items in the subscale or to the subscale total.

Though the means and standard deviations are roughly similar between the data in this study and the UK normative data, the distributions are not always so similar. This is most likely due to the small sample size in this study where less variation is acquired due to the smaller amount of data points. There seems to be more centring around the mean in this sample, which is reflected in the smaller standard deviations. It seems likely that more subjects would yield greater variation and thus more readily approximate the UK norms

Factor analysis 1

The third factor three emotional problem items and the conduct problem item *Fær oft skapofsaköst eða er uppstökkur*. It would seem then, that teachers thought this item to measure temper tantrums in an emotional sense, rather than an issue of conduct. This isn't entirely unreasonable since temper tantrums can quite easily be seen as being an emotional response**,** rather than poor conduct.

The semantic content of the items in factor 4 *Hjálpsamur/söm ef einhver meiðir sig, er í uppnámi eða líður illa; Góð/ur við yngri börn; Býðst oft til að hjálpa öðrum (foreldrum, kennurum, öðrum börnum)* concerns being helpful and being kind to younger children. These were all items from the prosocial subscale (and were the only items in this factor) but the other two items had their highest correlations on other factors. The prosocial items in factor 4 seemed to have more in common than the other two items. It is possible that teachers considered the item *Góð/ur við yngri börn* to assess helping younger children. The factor could thus be named Helpfulness.

The fifth factor consisted of 3 conduct problem items: *Flýgst oft á eða leggur börn í einelti; Almennt hlýðin/n, gerir yfirleitt eins og fullorðnir óska* and *Lýgur oft eða svindlar*. Some teachers had expressed concern over the first item, worried that the behaviours it was meant to assess were wildly different, too different in fact, to be measured by a single item. They considered *flýgst oft á* to mean to play roughly with your companions in a friendly manner, which is very different from bullying indeed. However it did make a factor with two other conduct problems items, so it didn't prove entirely problematic.

The five items that were changed from previous translations: *Truflast auðveldlega, hann/hún á erfitt með að einbeita sér; Öðrum börnum líkar almennt vel við hann/hana; Fær oft skapofsaköst eða er uppstökkur; Á auðvelt með að deila með öðrum börnum (nammi, dót, blýöntum o.s.frv.)* and *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* had mixed successes. The first two items had their highest correlations with their corresponding factors and were clearly part of those factors. The other three items did not fare so well. *Fær oft skapofsaköst eða er uppstökkur* landed in an emotional symptoms factor, *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* made up its own factor and finally, the item *Á auðvelt með að deila með öðrum börnum (nammi,*

dót, blýöntum o.s.frv.) had a negative correlation with a factor that also had another negatively correlated prosocial item and a positively correlated conduct problem item. The items *Fær oft skapofsaköst eða er uppstökkur* and *Á auðvelt með að deila með öðrum börnum (nammi, dót, blýöntum o.s.frv.)* seem to need revision, as these items did not have any high correlations with other factors. Though the item *Á auðvelt með að deila með öðrum börnum (nammi, dót, blýöntum o.s.frv.)* had a factor correlation of 0.31 with the factor containing 3 prosocial items. It seems likely that this item belongs to that factor, instead of being negatively correlated with *Stelur heima, í skóla eða annars staðar*. The item *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* had a correlation with an emotional symptoms factor that was only 0.04 lower than its highest correlation. It's entirely possible that this item will correlate higher with this emotional symptoms factor in a larger sample.

Factor analysis 2

In the second factor analysis (where 5 factors were determined to be drawn) there were also some issues that need to be discussed. The first factor had all of the hyperactivity items as well as 1 conduct problem item (almennt hlýðin/n, gerir yfirleitt eins og fullorðnir óska) as well as an emotional problems item (Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika). It is easy to see how the conduct problems item could be included in this factor since a child who is constantly fidgeting and getting out of his seat to roam around the classroom or forgets to do his homework can quite easily be seen as disobedient. The fact that the emotional symptoms item showed up in this factor is most likely quite coincidental, as this item didn't really have a high correlation with any item, not even to its own subscale score.

Even though all of the hyperactivity items landed on one factor, it did have some extra items as well. The prosocial scale, however was perfectly replicated, that is, all of the items in the subscale had their highest correlations with this factor (factor 2) and no other items had their highest correlations on it. This was most likely due to the fact that when the number of possible factors was deliberately limited, it forced these items together, which the previous factor analysis did not.

The third factor contained 3 emotional items *Óttast margt, verður auðveldlega hrædd/ur; Oft óhamingjusamur/söm, langt niðri eða tárast* and *Áhyggjur af mörgu, virðist oft áhyggjufull/ur* and 3 conduct problems items *Lýgur oft eða svindlar; Fær oft skapofsaköst eða er uppstökkur* and *Flýgst oft á eða leggur í einelti*. Given the semantic meaning of these items it's not entirely unreasonable to assume them to be related. If *Flýgst á* is seen as "acting out" or playing a little too rough (as proposed by some teachers), then it is possible that lying, having temper tantrums and acting out may be seen as symptoms of emotional problems, rather than issues of conduct. Especially if the teachers disregarded the bullying aspect of that item. It is however also possible that these items only came together in a factor due to the limited number of possible factors. The same emotional items landed on a common factor as in the previous factor analysis, where *Fær oft skapofsaköst eða er uppstökkur* also had its highest correlation. Even though 3 conduct problem items made a factor in the first analysis, it may not have been distinctive enough to make a separate factor in this second analysis and simply "piggy-backed" on *Fær oft skapofsaköst eða er uppstökkur*'s correlation with a stronger factor.

Factor 4 in this analysis consisted of four peer problem items as well as the conduct problem item *Stelur oft heima, í skóla eða annars staðar.* This item did not have any high correlations with other factors, but there is no good reason to believe that these items measure some common dimension. Their semantic contents are quite unrelated, unless you see stealing as a symptom of peer rejection. However, if this were the case it seems more reasonable to assume that stealing would be a part of a greater constellation of problematic behaviour, namely those of the conduct problems scale. It's not really clear but this could very well be an artefact of the small sample size.

The fifth factor consisted of the emotional item *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* and the peer problem *Frekar einræn/n, leikur sér oft ein/n.* It does make some sense that these items would be related, since insecurity is not far behind loneliness. The emotional problem item did not have any correlations over 0.10 with any other factor, so it seems less reasonable that its failure to load on an emotional problems factor was due to the small sample size. The peer problems item had a correlation of 0.379 with the factor

that had the other 4 peer problem items in it. This correlation might well increase and the erroneous correlation might decrease with a larger sample.

The 8-factor analysis (the first factor analysis) did explain more of the variance than the 5-factor analysis. Which is fairly self-evident since it produced a greater number of factors, thus explaining more of the variance. Both factor analyses replicated the hyperactivity factor, though this factor was coupled with two extra items in the 5-factor analysis. The 5-factor analysis replicated the peer problems factor perfectly and the 8-factor analysis replicated the prosocial factor. The 8-factor analysis almost replicated the peer problems factor in factor 4, which contained 4 peer problem items and 1 conduct problem.

A factor analysis that limits the number of possible factors suffers the disadvantage of having to force items into a limited number of factors, whereas the non-limited factor analysis can create factors for wayward items. This of course has its pros and cons. The benefit of the non-limited one is that we can easily see which items do not correlate well into factors, the downside of this free-factor creation is that we can end up with factors that don't really make sense, such as factor 7 in the the 8-factor analysis. Factor 7 in that factor analysis contained two negatively correlated prosocial items and the conduct problem item *Stelur heima, í skóla eða annars staðar.* The benefit of the limited-factor analysis is that it *will* force factors together (which is also its downside), coupling lesser correlations together rather than producing more factors. This, however also means that we can end up with less nonsensical factors, such as factor 3 of the 5-factor analysis, which had en equal mix of emotional problem items and conduct problems.

The reason for presenting both factor analyses in such detail is that none of them replicated the original five factor structure proposed by Goodman in 1994 (Goodman, 1994). Nor did one factor analysis seem better than the other. They both had their strengths and weaknesses, but neither was satisfactory. This could (again) simply be due to the small sample size but other reasons are discussed in the next section where the most problematic items as well as the items that were altered in this translation are analysed

Internal consistencies and inter-item, item-total correlations

As can be seen in table 5, the inter-scale correlations were not very high**.** This indicates that the subscales are measuring separate but slightly similar constructs. If the inter-scale correlations were perfect (i.e. r = 1) then they would measure only one construct, however, if we had correlations of 0, then they would not measure anything similar at all, which would be useless to us in this questionnaire since it would become much too unspecific.

What we have here is an indication that the questionnaire is measuring five distinct dimensions. Four of which are somewhat related but distinct (i.e. difficulties of different kinds) and one is separated from the rest (prosocial behaviours).

The internal consistencies, or the alphas, were generally satisfactory. However, the emotional symptoms scales alpha was lower than one would want and can be a cause of concern. The small sample size does not seem a likely culprit of this low internal consistency, since the other subscales have fairly high alphas. It seems more plausible that this is due to some peculiarities in the items themselves.

The fact that the prosocial correlations are quite low makes perfect sense since these items are not included in the total difficulties score. The fact that they are negative is due to the fact that they measure strengths, not difficulties. It may seem a bit silly to analyse these prosocial correlations, but the fact is that were these correlations not negative, it would mean that the data indicate a correlation between prosocial behaviours and difficulties. Goodman (1997) proclaims that the absence of prosocial behaviours is conceptually different from the presence of psychological difficulties, which is why these items are not scored in the reverse direction and incorporated in the total difficulties score (Goodman, 1997). But because of the way the items are scored (see *Scoring the SDQ* in introduction) it would be difficult to imagine a strong positive correlation between the total difficulties and prosocial scores. Even though it is far from impossible to display prosocial behaviours when one suffers from psychological difficulties, it seems that it would be less common in those suffering from peer problems, emotional problems and conduct problems especially.

The altered items

The item *Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust* from the emotional symptoms scale was one of the items that had been changed in this translation. It had 4 correlations under 0.3, though one of them was with the problematic item *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika*. So that particular correlation is most likely due to *Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika*'s problematic nature. Its weak inter-item correlations and weak correlation with total difficulties may explain why it failed to correlate with an emotional-problems factor in the factor analyses. In the first factor analysis it made up its own factor (being the one solitary item in that factor) and in the second it made up a factor with the item *Frekar einræn/n, leikur sér oft ein/n* - a peer problems item. Its correlations in the factor analyses were 0.82 and 0.75 in the first one and in the second one (in which 5 factor were determined to be drawn) respectively, having no correlations with other factors of that magnitude. The fact that it did not have any high correlations with other factors makes it seem less likely to be an effect of the small sample size, it would rather seem to need revision. Especially when an examination of the frequencies of this item revealed that a full 22.5% of endorsements on this item were for a score of 1 (*Að nokkru rétt*), and 3.8% were for a score of 2 (*Örugglega* rétt). Given these endorsements we see that quite a few children were rated as having at least some insecurity issues. Thus the item's failure to correlate into an emotional-problems factor does not seem to be due to any properties of the endorsement rates as such. It must have been seen as conceptually different from other emotional difficulties by the raters.

The item "*Fær oft skapofsaköst eða er uppstökkur*" was an item that had been altered for this edition of the Icelandic teacher-form SDQ. In both the 5-factor- and the 8-factor analysis did it correlate on a factor that contained the emotional symptoms items *Óttast margt; Óhamingjusamur* and *Áhyggjur*. Only in the 5-factor structure did *Skapofsaköst* correlate in to a factor which contained any conduct problem items, but this factor also contained the same emotional symptoms as the 8-factor analysis. The fact that this item correlated with the same emotional items in both factor analyses indicates that it was not actually assessing conduct problems, but rather emotional problems. It did have correlations with other conduct subscale

items and a satisfactory corrected item-total correlation of 0.54, but as the factor analyses showed, it seemed to have more in common with emotional items. However, as mentioned previously, some teachers had expressed concerns over the semantic duality of the item. It is in fact entirely possible that this item formed a factor with emotional symptoms simply because of the weak inter-item correlations of the emotional symptoms subscale in general and the perceived dualistic nature of *Flýgst oft á eða leggur börn í einelti*, and not because of a shared dimensional assessment.

The hyperactivity item that had been changed in this translation *Truflast auðveldlega, hann/hún á erfitt með að einbeita sér* had satisfactory item-total, inter-item and total difficulties correlations. As with all the hyperactivity items it consistently formed one cohesive factor. It was among the most successful items in the questionnaire, so this item clearly does not need revision

Another item that did not seem to need revision was *Öðrum börnum líkar almennt vel við hann/hana*, a peer problems item. It had satisfactory correlations and it successfully correlated with its predetermined factor in both factor analyses. Thus the changes made to this item seem sufficient for this age group. The last of the changed items (*Á auðvelt með að deila með öðrum börnum*) was not a successful item. In the first factor analysis its highest correlation was a negative correlation to a factor that also contained another negatively correlated prosocial item (*Tekur tillit til tilfinningar annarra*) and a positively correlated conduct problems item (*Stelur heima, í skóla eða annars staðar*). However, in the factor analysis in which five factors were drawn, it had its highest correlation with a perfectly replicated prosocial factor. These results are initially perplexing, but the fact that the five-factor analysis succeeded in replicating the factor where the eight-factor analysis failed is perhaps not so mysterious. The eight-factor analysis must have put the negatively charged correlations with the ever wayward item *Stelur heima, í skóla eða annars staðar* in an attempt to make a factor. However, *if* the absence of prosocial behaviours is distinctly different from difficulties, then a negative prosocial and positive conduct problems factor does not make sense. It is much more reasonable to assume that this effect was due to the nature of factor analyses where there are no restrictions on how many factors can be drawn. It does not seem like this item will need revision, as it did correlate well into its predetermined factor in the 5-factor analysis. Its correlation, and the correlation of *Tekur tillit til tilfinningar annara,* in the 8-factor analysis were most

likely due to the problematic nature of *Stelur heima, í skóla eða annars staðar* and their relatively low inter-item correlations within the prosocial subscale (*Tillit*'s mean inter-item correlation was 0.26), rather than these three items assessing a common dimension.

Other problematic items

The item Stelur heima, í skóla eða annars staðar in particular was not successful at all. It had very low and negative inter-item correlations, an unsatisfactory item-total correlation as well as a very low correlation with the total difficulties score, nor did it correlate well with its intended factor in either analysis. It did not seem to be seen as a conduct problem, but instead, as the factor analyses seem to imply, as an issue relating to either peer problems or absence of prosocial behaviours. It is initially puzzling that it correlated so poorly with the total difficulties score, until you examine the frequency distribution of this particular item. Only two subjects received a score of 1 (Að nokkru rétt) on the stealing item, whereas the rest received a 0 (Ekki rétt). This is most likely an artefact of the small sample size. Even if we can assume that the proportions of responses would remain the same, we would have more endorsements of 1's and possibly 2's (Örugglega rétt). We would then have more instances of high scores on this item together with high scores on other items in the total difficulties score. This is purely speculation of course, but it seems more reasonable to assume that the peculiar properties of this item are due to data-abnormalities rather than some semantic similarities between this item and peer problems or prosocial items.

Limitations of the study and directions for further research

The most obvious limitation of this study was the small sample size. There was a very small participation rate (as described in the Method section). The participation rate was most likely diminished by the fact that data collection could not start until March. This late in the semester there are already many studies being conducted in the schools and most head masters were hesitant to burden their teachers with more work. Also, most teachers were hesitant to take part when the head master had given the experimenter permission to approach the teachers.

Another limitation was the narrow scope of the study. This of course is only a bachelor's thesis, but it is a part of a larger research project in which other forms of the questionnaire and other age groups are also examined. If these were coordinated to target the same children then ratings could be compared in order to examine rater characteristics and determine if any items consistently differed between rater types in a way that was not expected. This study merely concerns the psychometric properties of the teacher form SDQ for the ages 6-10, which is more than enough for a bachelor's thesis. But if another student had the parent form for this age group then they could organize to have the same children rated and comparisons could be made (though not necessarily by these students). This would utilize the data collection to a much greater extent. In addition, if the numbered lists of the participants (see Method section) were kept, then the test-retest reliability could be examined.

The results of this study were unsatisfactory. The failure of the five-factor structure to replicate, the sometimes low inter-item correlations and items that didn't seem to belong anywhere would of course not bring a swift conclusion to the aforementioned research-endeavour. However, were the sample size large enough then it would seem likelier that good results could be reached and thus proper comparisons could be made.

Some items seemed to stand out. Some had low inter-item and/or item-total correlations, some formed their own factors and others landed on factors that didn't make much sense. Some items seemed to assess completely different constructs than they were designed to. These items would have to be revised before going onwards to the large-scale study, especially if they were also found to be problematic in the other studies.

Conclusions

The subscales internal consistencies were mostly adequate, indicating that for the most part, items within each subscale were measuring similar constructs. This, however, was not found by either factor analysis. There were items correlating with factors that were quite contrary to the original five factor structure, items that made up their own, separate factors and factors that didn't make much sense.

Based on the results of this study, it doesn't look like this translation of the Icelandic teacher-form SDQ can be used for this age group (6-10 years of age). However, it must be stated that the sample was much too small in order to draw any definite conclusions. In fact, Einar Guðmundsson and Árni Kristjánsson (2005) suggest that we should not even conduct factor analyses on samples smaller than 100 participants, since the results from such small samples might not be replicable with other samples. Even so, the data at hand must be discussed since they might carry some significance for further research.

In spite of the relatively good internal consistencies, the inter-item correlations were generally somewhat poor and in the case of the conduct problems scale, quite poor. The item *Stelur heima, í skóla eða annars staðar* was by far the least successful item in the questionnaire. Stealing in this age group would be rather rare, as evidenced by the fact that only two subjects received a rating of 1 and no subject received a 2 on this item. With only two data entries over 0, it would be quite hard to make any decent correlations. Despite its poor performance, this item probably does not need revision, since it is basically a word-by-word translation of the English item to Icelandic, and there is little room for misinterpretation. Its small correlations were most likely caused by the small sample size.

Another problematic item in the conduct scale was "*Fær oft skapofsaköst eða er uppstökkur*". Seeing as how this item seemed to assess emotional symptoms more than conduct problems, this item would need to be revised again. So possible reasons for rewording it would have to be two-fold. First, it must be seen as assessing one singular type of problem, and secondly, it must assess conduct problems and not emotional ones. If *Flýgst á* is considered playing roughly, then it is far from the act of bullying. The original English wording of the item is: "Often fights with other children or bullies them", which is not far from the act of bullying. To play

roughly, on the other hand, does not have the same negative connotations as fighting. Something along the lines of "Réðst oft á eða leggur börn í einelti", or "Lendur oft í slag eða leggur börn í einelti", might yield better results, something that emphasizes the aggressive aspect of the first part of the item.

Two emotional symptoms items were very unsatisfactory. The first one, "Kvartar oft um höfuðverk, magaverk eða flökturleika oft um höfuðverk, magaverk eða flökturleika" , consistently failed to be useful. In the factor analyses it either formed its own factor (8-factor analysis) or correlated with an unpredicted factor (5-factor analysis). It is possible that the item was misinterpreted to mean "often *has* headaches (etc.)". It may be proper to emphasize the word "Kvartar oft um höfuðverk, magaverk eða flökturleika" by italicizing it so it will be written as: "*Kvartar oft um höfuðverk, magaverk eða flökturleika* oft um höfuðverk, magaverk eða flökturleika" or perhaps put it in bold face. This is purely speculation. There is no evidence that this item was misinterpreted in this way but no other explanation comes to mind, since there was little indication that this item had much in common with any other item.

The other emotional item which was problematic was "Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega sjálfstraust". This was one of the items that had been altered for this edition. It didn't have much in the way of proper correlations with other items in the emotional symptoms subscale, nor did it do well in the factor analyses. The frequency distribution revealed that a full 22.5% received a rating of 1 and 3.8% received a 2. So obviously some children were rated as being somewhat insecure. However, this item asks teachers to rate whether children cling to their *parents* in new situations. This would mean that teachers would have to assess something that is most likely put of their field of experience. Granted, parents drop of their children at school, which is a new (and sometimes scary) experience the first few days. But after that it becomes a part of the daily routine and is thus not a new situation. It does not seem very reasonable to expect that teachers would have much experience of how children behave towards their parents in new situations, since they most likely do not have much access to this. This item might fare better if the reference to their parents were removed. So that the item might be written as such: "Óörugg/ur við ókunnar aðstæður, missir auðveldlega sjálfstraust.". There is a reference to being *clingy* in the English version of this item, but it does not refer

specifically to the parents. So *clingy* in the English item could easily refer to the the children clinging on to the teachers or perhaps other children or even a precious toy of some kind, when they go on field-trips for instance.

The failure of the factor analyses to replicate the five factor structure, with only the hyperactivity factor being reproduced in both analyses indicate that only one of the five dimensions were tapped in this sample. Once again the small sample size must be mentioned as a likely cause. The fact that the hyperactivity scale was replicated is not surprising since hyperactive symptoms are quite common in children and they're easily seen as they are completely overt. So this would suffer much less from a small sample size due to its commonness. Conduct problems (as formulated in the items) are of course also very much overt, but much less common than hyperactivity. So they would suffer more from a small sample size. Of course, this being a behavioural screening questionnaire, it does assess behaviours, or at least symptoms that can be seen. However the emotional problems scale does require the rater to "go inside the childs head" to some extent, having to assess and infer if the child is worried, anxious, insecure etc., which should be less reliable than observing behaviours such as hyperactivity.

All in all, based on these results, the Icelandic teacher form SDQ cannot be used for 6-10 year-olds. It didn't replicate the five factor structure, nor did a useful alternative factor structure emerge. Most of the cut-offs were nowhere near Goodman's (1997) suggested cut-off scores and the use of the 90th percentile to indicate abnormality didn't seem useful - except for with the hyperactivity scale, which was very close to Goodman's (1997) suggestions and the prosocial scale, which was somewhat close.

References

Achenbach, T. M. (2001). Challenges and benefits of assessment, diagnosis, and taxonomy for clinical practice and research. *Australian and New Zealand Journal of Psychiatry, 35,* 263- 271.

Achenbach, T. M., McConaughy, S. H. & Howell, C. T. (1987). Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Informant Correlations for Situational Specificity. *Psychological Bulletin, 101,* 213-232.

Agnes Huld Hrafnsdóttir. (2005). Athugun á próffræðilegum eiginleikum Spurninga um styrk og vanda í hópi 5 ára barna á Íslandi. *Sálfræðiritið - Tímarit Sálfræðingafélags Íslands*, *10,* 71- 82.

Auður Magnúsdóttir & Berglind Sveinbjörnsdóttir (2004). *Próffræðilegir eiginlekar Íslenskrar útgáfu Foreldra og Kennaramatskvarða SDQ.* Unpublished bachelors thesis: Háskóli Íslands

Avenevioli, S., Knight, E., Kessler, R. C. & Merikangas, K. R. (2008). Epidemiology of depression in children and adolescents. In Abela, J. Z .R & Hankin, B. L. (editors). *Handbook of depression in children and adolescents (pp.*6-32). New York: Guilford Press.

Becker, A., Steinhausen, H. C., Baldursson, G., Dalsgaard, S., Lorenzo, M. J., Raltson, S. J., Döpfner, M. & Rothenberger. A. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Child & Adolescent Psychiatry, 15*, 56-62

Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: Do the self-reports tell us more than ratings by adult informants? *European Child & Adolescent Psychiatry, 13*, 17-24.

Bongers, I.L., Koot,H.M., van der Ende, J., & Verhulst, F.C. (2003).The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology*, *112,* 179–192.

Cole, M., Cole, S. R. & Lightfoot, C. (2005). *The Development of Children* (5.ed). New York: Worth Publishers.

Dickey, W. C. & Blumberg, S. J. (2001). Revisiting the Factor Structure of the Strengths and Difficulties Questionnaire: United States, 2001 [Abstract]. *Journal of the American Academy of Child & Adolescent Psychiatry, 43,* 1159-1167. Downloaded 28 April 2011 from http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B987N-4XKH917G&_user=5915660&_coverDate=09%2F30%2F2004&_rdoc=1&_fmt=high&_orig=gateway&_origin=gateway&_sort=d&_docanchor=&view=c&_rerunOrigin=scholar.google&_acct=C000068853&_version=1&_urlVersion=0&_userid=5915660&md5=3d5cd13917864f3c55b080db71b18a60&searchtype=a

Einar Gúðmundsson & Árni Kristjánsson. (2005). *Gagnavinnsla í SPSS.* Reykjavík: Háskólaútgáfan.

Furr, R. M. & Bacharach, V. R. (2008). *Psychometrics: An introduction.* Tousand Oaks: Sage publications Inc.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child psychology and Psychiatry, 38,* 581-586.

Goodman, R. (1999). The Extended Version of the Strengths and Difficulties Questionnaire as a Guide to Child Psychiatric Caseness and Consequent Burden. *Journal of Child Psychology and Psychiatry, 40,* 791-799

Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *British Journal of psychiatry, 177,* 534-539.

Goodman, R. & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child behavior Checklist: Is Small Beautiful? *Journal of Abnormal Child Psychology, 27,* 17-24.

Harpa Hrund Berndsen. (2005). *Próffræðilegir eiginleikar Íslenskrar þýðingar spurningalistans um styrk og vanda (SDQ).* Unpublished bachelors thesis: Háskóli Íslands.

Koskelainen, M., Sourander, A. & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child & Adolescent Psychiatry, 10,* 180-185.

Lai, K. Y. C., Luk, E. S. L., Leung, P. W. L., Wong, A. S. Y., Law, L & Ho, K. (2010). Validation of the Chinese version of the strenths and difficulties quesionnaire in Hong Kong. *Social Psychiatry and Psychiatric Epidemiology, 45,* 1179-1186.

Mash, E. J. & Wolfe, D. A (2010). *Abnormal Child Psychology* (4.ed). Belmont: Wadsworth.

Muris, P., Meesters, C. & van den Berg, F. (2003) The Strengths and Difficulties Questionnaire (SDQ). Further evidence for its reliability and validity in a community sample of Dutch children and adolescents. *European Child & Adolescent Psychiatry, 12,* 1-8.

Obel, C., Heiervang, E., Rodriguez, A., Heyerdahl, S., Smedje, H., Sourander, A., Guðmundsson, Ó. Ó., Chlench-Aas, J., Christensen, E., Heian, F., Mathiesen, K. S., Magnússon, P., Njarðvík, U., Koskelainen, M., Rønning, J. A., Stormark, K. M. & Olsen, J. (2004).The Strengths and Diccifulties questionnaire in the Nordic countries. *European Child & Adolescent Psychiatry, 13*, 32-39.

Reber, A. S., Allen, R. & Reber, E. S. (2009). *Penguin Dictionary of Psychology* (4.ed). London: Penguin Group.

Ross, A. O. (1987). *Personality: The scientific study of complex human behavior* (2. chapter). New York: Holt, Reinhart & Wilson.

Smith, S. R. (2007). Making Sense of Multiple Informants in Child and Adolescent Psychopathology. A guide for clinicians. *Journal of Psychoeducational Assessment*, *25,* 139-149.

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A. & Janssens, J. M. A. M. (2010). Psychometric Properties of the Parent and Teacher Versions of the Strengths and Difficulties Questionnaire for 4-12 year-olds: A Review. *Clinical Child and Family Psychology Review, 13,* 254-274.

The National Child Traumatic Stress Network. *Strengths and Difficulties Questionnaire-Teacher Report*. Downloaded 2 Febuary 2011 at http://www.nctsn.org/sites/default/files/assets/pdfs/measures/SDQ-teacher.pdf

van Widenfelt, B. M., Goedhart, A. W., Treffers, P. D. A. & Goodman, R. (2003). Dutch version of the Strengths and Difficulties Questionnaire (SDQ). *European Child & Adolescent Psychiatry, 12,* 281-289.

Youth in Mind. Information for researchers and professionals about the Strengths and Difficulties Questionnaires. http://sdqinfo.org.

Appendices

Appendix A, The current translation of the Icelandic SDQ teacher form and impact supplement.

Appendix B, The e-mail provided to teachers to send to parents or guardians to obtain their informed constent

Appendix A

Spurningar um styrk og vanda (SDQ-Ice) K 4-16

Svarið hverri fullyrðingu, með því að merkja í einn reit: Ekki rétt, Að nokkru rétt eða Örugglega rétt. Þið eruð beðin að merkja við allar fullyrðingarnar, jafnvel þótt þið séuð ekki alveg viss, eða þær sýnast heimskulegar. Svarið með tilliti til atferlis barnsins síðustu sex mánuði eða yfirstandandi skólaár.

Kyn barns: _____Stúlka_____Drengur

Aldur barns:_____ára

	Ekki rétt	**Að nokkru rétt**	**Örugglega rétt**
Tekur tillit til tilfinninga annarra	☐	☐	☐
Eirðarlaus, ofvirk/ur, getur ekki verið kyrr lengi	☐	☐	☐
Kvartar oft um höfuðverk, magaverk eða flökurleika	☐	☐	☐
Á auðvelt með að deila með öðrum börnum (nammi, dóti, blýöntum o.s.frv.)	☐	☐	☐
Fær oft skapofsaköst eða er uppstökkur	☐	☐	☐
Frekar einræn/n, leikur sér oft ein/n	☐	☐	☐
Almennt hlýðin/n, gerir yfirleitt eins og fullorðnir óska	☐	☐	☐
Áhyggjur af mörgu, virðist oft áhyggjufull/ur	☐	☐	☐
Hjálpsamur/söm ef einhver meiðir sig, er í uppnámi eða líður illa	☐	☐	☐
Stöðugt með fikt eða á iði	☐	☐	☐
Á að minnsta kosti einn góðan vin	☐	☐	☐
Flýgst oft á eða leggur börn í einelti	☐	☐	☐
Oft óhamingjusamur/söm, langt niðri eða tárast	☐	☐	☐
Öðrum börnum líkar almennt vel við hann/hana	☐	☐	☐
Truflast auðveldlega, hann/hún á erfitt með að einbeita sér	☐	☐	☐
Óörugg/ur, hangir í foreldrum við ókunnar aðstæður, missir auðveldlega Sjálfstraust	☐	☐	☐
Góð/ur við yngri börn	☐	☐	☐
Lýgur oft eða svindlar	☐	☐	☐
Verður fyrir stríðni eða einelti af hálfu annarra barna	☐	☐	☐
Býðst oft til að hjálpa öðrum (foreldrum, kennurum, öðrum börnum)	☐	☐	☐
Hugsar áður en hann/hún framkvæmir	☐	☐	☐
Stelur heima, í skóla eða annars staðar	☐	☐	☐
Semur betur við fullorðna en önnur börn	☐	☐	☐
Óttast margt, verður auðveldlega hrædd/ur	☐	☐	☐
Fylgir verkefnum eftir til enda, heldur góðri athygli	☐	☐	☐

Gerið svo vel að fletta – það eru nokkrar spurningar á næstu síðu

Frekari athugasemdir eða áhyggjur sem þið kynnuð að hafa:
Almennt séð, telur þú að barnið eigi við erfiðleika að stríða á einu eða fleirum af eftirtöldum sviðum: Tilfinningar, einbeiting, hegðun eða samspil við aðra?

Nei	Já væga erfiðleika	Já greinilega erfiðleika	Já alvarlega erfiðleika
☐	☐	☐	☐

Ef svarið var „Já“ gerið þá svo vel að svara eftirfarandi spurningum um þessa erfiðleika?

Hve lengi hafa þessir erfiðleikar verið til staðar?

Minna en mánuð	1-5 mánuði	6-12 mánuði	Meira en ár
☐	☐	☐	☐

Valda þessir erfiðleikar barninu þínu hugarangri eða vanlíðan?

Alls ekki	Lítils háttar	Þó nokkuð	Mjög mikið
☐	☐	☐	☐

Trufla þessir erfiðleikar daglegt líf barnsins á eftirfarandi sviðum:

	Alls ekki	Lítils háttar	Þó nokkuð	Mjög mikið
SAMSKIPTI VIÐ JAFNALDRA	☐	☐	☐	☐
NÁM Í SKÓLANUM	☐	☐	☐	☐

Eru þessir erfiðleikar barnsins íþyngjandi fyrir þig eða bekkinn í heild?

Alls ekki	Lítils háttar	Þó nokkuð	Mjög mikið
☐	☐	☐	☐

Dagsetning .....

Kærar þakkir fyrir hjálpina

Appendix B

Kæru foreldrar og forráðamenn

Barnið þitt hefur lent í úrtaki fyrir rannsókn á spurningalista sem metur hegðun og líðan barna og er nú verið að prófa íslenska þýðingu á breskri útgáfu hans (*Strenghts and Difficulties Questionnaire –SDQ* – á íslensku *Spurningar um styrk og vanda*). Nokkur börn voru valin tilviljanakennt úr bekk barnsins þíns og kennarinn hefur verið beðinn um að fylla listann út með barnið þitt í huga.

SDQ listinn hefur komið að gagni við að skima fyrir hegðunar – og tilfinningaerfiðleika barna. Hann hefur verið notaður um árabil við greiningu íslenskra barna (t.d. á BUGL, Greiningar- og ráðgjafarstöð ríkisins, í skólakerfinu og heilsugæslu). Þýðing listans hefur hins vegar verið gagnrýnd og er tilgangur rannsóknarinnar að skoða nýja og endurbætta útgáfu listans.

Kennarinn mun fylla út listann með barnið þitt í huga. Ekki er beðið um neinar persónuupplýsingar þar sem listinn verður hvorki merktur forráðamanni, kennara né barni. Svör eru því ekki rekjanleg til einstakra þátttakenda.

Hér með er óskað eftir leyfi þínu fyrir að listinn sé fylltur út með þitt barn í huga. Þér ber að sjálfsögðu *engin skylda* til að leyfa kennaranum að taka þátt í þessari athugun.

Ef þú ert mótfallin(n) því að listinn sé fylltur út fyrir þitt barn ertu vinsamlegast beðin(n) um að láta kennarann vita sem fyrst, fyrir _____.

Berist engar athugasemdir frá þér fyrir þann tíma verður litið svo á að þú gefir samþykki fyrir þátttöku kennarans í þessari athugun.

Kærar þakkir,

Dr. Fanney Þórsdóttir, lektor við Sálfræðideild Háskóla Íslands
Dr. Urður Njarðvík, lektor við Sálfræðideild Háskóla Íslands

Ef spurningar vakna vegna hegðunar eða líðan barnsins þíns við þátttöku í þessari rannsókn er þér velkomið að hafa samband við Dr. Urði Njarðvík, barnasálfræðing og lektor við sálfræðideild HÍ. (urdurn@hi.is, sími: 525-5957)