Psychometric Validity and Reliability of the Social Skills Improvement System-Rating

Scales (SSIS-RS)

Thor Klaussen*, Lene-Mari P. Rasmussen*, Børge Strømgren**, and Frode Svartdal*

*University of Tromsø and **Oslo and Akershus University College

Corresponding author: F. Svartdal, Department of psychology, University of Tromsø, 9037

Tromsø, Norway. E.mail: frode.svartdal@uit.no

Abstract

The *Social Skills Improvement System-Rating Scales (SSIS-RS)* assesses social skills and problem behavior in children and adolescents from three perspectives (teacher, parent, and student). The SSIS-RS is a revised version of the *Social Skills Rating System (SSRS)*. A Norwegian translation of the SSRS has been validated, but this is not the case for the Norwegian translation of the SSIS-RS. The aim of this study was to compare the former translation of the SSRS with a translation of the SSIS-RS in samples of children (aged 8-12 and 13-16) as well as with their parents and teachers (total $N = 599$). The results indicated moderate to strong relations between the common subscales across all forms of the two instruments and acceptable to excellent internal consistency across all common subscales. We conclude that the SSIS-RS is a promising instrument for measuring social skills and problem behavior among children and adolescents in Norway.

Keywords: Social Skills Rating Scales (SSRS), Social Skills Improvement System – Rating Scales (SSIS-RS), validity, reliability

Psychometric Validity and Reliability of the Social Skills Improvement System-

Rating Scales (SSIS-RS)

Empirical studies highlight the importance of social skills and social competence among children and youths (Elliot & Gresham, 1987; Langeveld, Gundersen, & Svartdal, 2011; Ogden, 2011). Social competence among children is reflected in the context of a satisfactory academic performance and positive peer relationship (Gresham, Elliot, Cook, Vance, & Kettler, 2010a), as well as positive relations to adults (Ogden, 2011). Children with social competence deficits will often show difficulties in the development and maintenance of interpersonal relationship, display a lack of pro-social behavior patterns and poor academic achievement is often present. The negative consequences associated with social competence deficits have proven to influence a child's life throughout childhood, and may continue into adulthood, affecting domains like education and psychosocial function (Kupersmidt, Coie & Dodge, 1990; Newcomb, Bukowski, & Pattee, 1993). Research indicates that youngsters with interpersonal problems and social competence deficits are at risk for developing many damaging life-consequences, e.g. school dropout, juvenile delinquency, adulthood psychopathology, depression, and suicide (Cowen, Pederson, Babigian, Izzo, & Trost, 1973; Kohn & Clausen, 1955; Kupersmidt, Coie, & Dodge, 1990; Parker & Asher, 1987).

Social skills constitute an essential part of social competence, and deficits in this area are therefore present as a characteristic in many disabilities, for instance emotional and behavioral disorders (Gresham, Cook, Crews, & Kern, 2004; Maag, 2005), specific learning disabilities (Gresham, 1992), attention deficit/hyperactivity disorder (Smith, Barkley, & Shapiro, 2007), conduct disorder (Dodge & Pettit, 2003), and mild mental retardation (Gresham & Reschly, 1987). Also, social competence deficits and difficulties in interpersonal relationships are part of several diagnostic criteria specified in the *Diagnostic and Statistical*

Manual of Mental Disorders (4ᵗʰ ed., text rev.; DSM-IV TR; American Psychiatric Association, 2000).

It is evident; therefore, that good measurement of social competence and problem behavior is important. There are many available instruments claiming to measure these constructs (Svartdal & Klaussen, 2013), some using only one informant and others using multiple informants. Multi-informant instruments with versions for child, parent and teacher are less common, but probably preferable because each rater has access to unique information about the individual being rated (Gresham et al., 2010a; Renk & Phares, 2004).

One such multi-rater instrument to assess children`s social skills and problem behavior is the *Social Skills Improvement System* (*SSIS-RS;* Gresham & Elliot, 2008). The SSIS-RS is a revised version of the *Social Skills Rating System* (*SSRS*; Gresham & Elliott, 1990), and these instruments have much the same structure. The SSRS was designed over 20 years ago for the purpose of being a brief and comprehensive tool in identifying children at risk for social behavior difficulties, and selecting target behaviors for interventions (Gresham & Elliot, 1990). The SSRS is among the most frequently used instruments for measuring children's (age 3-18) pro-social behavior and problem behavior in schools in the United States and other countries. The instrument is also widely used in research with children, and is one of the most popular rating scales measured in citations (Crowe, Beauchamp, Catroppa, & Anderson, 2011). A number of studies have supported the validity of both instruments (Walthall, Konold & Pianta, 2005; Frey, Elliott, & Gresham, 2011; Gresham & Elliot, 1990, 2008; Humphrey et al., 2011), and recently the SSIS-RS was validated against SSRS (Gresham, Elliott, Vance, & Cook, 2011).

The SSRS was translated into several other languages including Norwegian, Spanish, Portuguese, Hindi, Dutch, Iranian, Slovakian, German, Russian, and Korean (Gresham et al., 2011). Several studies support the validity and reliability of the translated versions of SSRS

(Juardo, Cumba-Aviles, Collazo, & Matos, 2006; Ogden, 2003; Shahim, 2001; Van der Oord et al., 2005; Vasil'ová & Baumgartner, 2004). In Norway, the instrument was translated and validated ten years ago (Ogden, 2003). It was subsequently used to measure treatment effect of Aggression Replacement Training (ART), Social Perception Training (SPT), Multisystemic Treatment (MST) and Parent Management Training Oreagon model (PMTO) (Gundersen & Svartdal, 2006; Langeveld et al., 2011; Ogden & Hagen, 2008; Ogden & Halliday-Boykins, 2004), and other research purposes (Sørlie, Hagen, & Ogden, 2008), including master theses (Sæstad & Kyrrestad, 2007; Nyquist & Rasch-Olsen, 2011; Eliassen, 2012). The SSRS have also been used to validate other rating scales (Bjørnbekk & Howard, 2012).

Over the years, the SSRS have met some concerns regarding its theoretical framework and technical features, as well as out dated items. Consequently, a revised version − the SSIS-RS − was developed to address these concerns. In Norway the SSIS-RS have been translated (Strømgren, 2012), but not yet validated. To secure validation, a simple comparison of the Norwegian versions of the SSIS-RS and SSRS could be performed by examining the Pearson *r* correlations between comparable subscales of the individual tests. This was the procedure selected when Gresham and colleagues (Gresham et al., 2011) validated the SSIS-RS. However, additional issues arise when a translated version is to be validated.

First, the validation of SSIS-RS depends on the quality of the original validation of the SSRS (Ogden, 2003). However, this validation was probably not optimal. It focused only on the teacher ratings for secondary aged students, and the scale was not validated against other established behavioral rating scales. Furthermore, the Norwegian SSRS version had a few modifications from the original SSRS (e.g. an increase from a 3-point to a 4-point scale in assessment of social skills; a decrease from a 5-point to a 4-point scale when assessing academic competence; the parent form was reduced from 40 to 23 items). Secondly, issues towards cross-cultural equivalence should be considered since this is a translated instrument

adapted from another country. Although equivalence is not a primary objective in this study, some empirical investigations of cross-cultural equivalence will be presented.

The purpose of the present study was to compare the Norwegian translation of the SSIS-RS to the translated SSRS. To clarify the issues involved in this process we briefly discuss the procedure Gresham et al., (2011) selected to validate the SSIS-RS. Then we discuss issues involved in doing a corresponding validation for the Norwegian SSIS-RS version, including issues related to differences in culture.

SSRS and SSIS-RS Description. The SSRS and SSIS have been described in earlier studies, both in terms of description (Scales, Subscales, Items) and psychometric properties (Albertus, Birkinbine, Lyon, & Naibi, 1996; Bramlett, Smith, & Edmond, 1994; Flanagan, Alfonso, Primavera, Povall, & Higgins, 1996; Frey, Elliott, & Gresham, 2011; Gresham & Elliot, 1990; Gresham & Elliot, 2008; Gresham et al., 2011; Manz, Fantuzzo, & McDermott, 1999; Merrell & Popinga, 1994; Ogden, 2003; Van der Oord et al., 2005; Whiteside, McCarthy, & Miller, 2007). Space does not allow a recapitulation of reliability and validity details, but a description of the main scale features of the SSRS and the SSIS-RS are provided in Table 1.

--- Table 1---

The US SSRS – SSIS-RS validation. There are several ways to assess the validity of an instrument, and one important type is construct validity. However, there is no easy way to directly assess construct validity, and therefore convergent and divergent/discriminant validity as indicated by correlations are good indicators of construct validity (Campell & Fiske, 1959). Gresham et al., (2011) applied this method indicating convergent and divergent relations by correlations between like-named total scales and subscales. In additon, Gresham et al., (2011) performed a detailed comparison of SSRS and SSIS-RS based on alpha estimates. They found, as excpected, evidence of convergent and divergent relations. Further, acceptable alpha

coefficients for all SSIS-RS scales, and comparisons to like-named scales on SSRS showed mostly significantly higher alphas for SSIS-RS scales. Therefore, based on Gresham et al., (2011), the main focus of this study is to present evidence of convergent and divergent validity based on Pearson r correlations between like-named total scales and subscales, and to compare alpha coefficients produced for like-named scales across SSIS-RS and SSRS.

Gresham et al., (2011) collected participants from several sites that were included in the SSIS-RS national norm sample and consisted of individuals that represented all major demographics. This included 221 elementary and secondary teacher ratings, 240 parent ratings for elementary and secondary school students, and 224 self-ratings from elementary and secondary school students (total $N = 665$). In the present study, participants were recruited from two towns in northern Norway (total $N = 599$). As the Norwegian population is a relative homogenous group, this is believed not to be critical regarding the representativeness of the sample to the general Norwegian population (Levinson, 1998; Statistics Norway, 2013).

The 2003 SSRS Norwegian validation. Whereas Gresham et al., (2011) based their validation on the original version of SSRS; the present study based the comparison on the Norwegian translation of the SSRS, in where the social skill scales are rated on a 4-point scale (Ogden, 2003). This change was made because Ogden (2003) performed a pilot study in which the teachers requested a 4-point scale due to the difficulty in differentiating with the 3-point scale. There is only one published study validating this Norwegian version (Ogden, 2003), in a sample consisting of secondary students (age 13-14) recruited from a Norwegian municipality ($N = 395$). Ogden reported that most students came from families with high education and income levels compared to Norwegian standards.

The results indicated internal consistency estimates ranging from .88 to .94, with a mean Cronbach`s alpha of .91 for both subscales and total scales. Additionally, further

indication of internal consistency was shown by significant positive inter-correlations ($p <$ .001) between social skills subscales, and between the subscales and social skills total scale score. The test-retest correlation was $r = .58$, demonstrating a relatively good stability measure over a period of 17 months.

Factor analysis supported the 3-factor structure reported by Gresham and Elliott, (1990), including similar factor loadings. Further, convergent validity was indicated by moderate positive correlations between both social skills scales and problem behavior scales across raters (teacher, parent, and students), and by positive correlations between social skills and academic competence, grade point average and peer nominations. In addition, divergent validity was indicated by negative correlations between teacher total social skills score and teacher problem behavior subscale scores. Additional evidence of construct validity was shown by significantly lower social skills estimates produced for student referred to special education, school psychologists, child and adolescent psychiatry, and child welfare, than students not referred ($p < .001$). Ogden (2003) concluded that SSRS teacher ratings could be used as an instrument for measuring social skills among Norwegian secondary students.

As mentioned, some limitations regarding the Ogden (2003) study should be noted. First, it only focused on the teacher ratings for secondary aged students. Secondly, the SSRS was not validated against other behavioral rating scales besides the other forms of SSRS (Parent and Student)[1]. Third, the SSRS version used in this study had a few modifications from the original SSRS. This included an increase from a 3-point to a 4-point scale when assessing social skills; a decrease from a 5-point to a 4-point scale when assessing academic competence; and the Parent form was reduced from 40 to 23 items for practical reasons.

Despite these issues, we believe that the Norwegian version of SSRS may be considered as an acceptable foundation for a validation study of the Norwegian version of

---

[1] There is a consistent finding in research that multi-informant rating scales used to evaluate children and adolescents show low inter-rater agreement among informants (Achenbach et al., 1987, Gresham et al., 2010a, Renk & Phares 2004). Therefore, this method of validation could be problematic.

SSIS-RS. This is based on (a) the theoretical basis and psychometrical properties of SSRS (Demaray et al., 1995; Gresham & Elliot, 1990; Gresham et al., 2011); (b) the results of the Norwegian validation study; (c) the frequent use of SSRS in research and assessment of school childrens social skills and problem behavior; (d) studies reporting validity and reliability of translated versions of SSRS in several languages, also showing equivalence, and thereby further evidencing generalized validity for SSRS (European Federation of Psychologist Associations, 2009). Additionally, validity for the Norwegian version of SSRS has been shown by comparisons with CADBI, CBCL, and HIT in an effect study of ART (Gundersen & Svartdal, 2006).

Cultural issues in validation. In cross-cultural research, equivalence is a fundamental methodical problem and should be considered when adopting rating scales from a different culture. From a measurement perspective, there are a number of different hierarchically linked types of equivalence that assume increasingly stronger level of measurement comparability across cultures (Kankaras & Morse, 2010, p. 124). Culture differences between American and Norwegian cultures clearly exist, and these might influence the utility of the instrument across the two cultures. Consequently, estimates of validity and reliability must be separately established within each culture. Cross-cultural equivalence was not a primary focus in this study. However, issues regarding linguistic, conceptual and metric equivalence will be examined and discussed to some extent.

Linguistic (or translation) equivalence emphasis the language used in research, including questionnaires, and a translation should be smooth and natural sounding in the second language (Gudykunst, 2002). Back-translation is often used, involving one bilingual to translate the questionnaire into the second language (forward translation) then another bilingual translates back to the first language (back-translation) (Gudykunst, 2002).

Back-translation method may also be used to establish conceptual equivalence, the tests ability to measure the same construct and meaning in both or all cultures (Geisinger, 2003). Support for conceptual equivalence can be indicated by evidence of convergent and divergent relationships.

Further, metric equivalence is supported when similar psychometric properties are produced by the instrument in both cultures and languages (Geisinger, 2003). Therefore, in accordance with supporting metric equivalence, comparisons will be made of internal consistency estimates calculated for the Norwegian and US version of SSIS-RS. Similar levels of internal consistency would evidence metric equivalence of the instrument across the two cultures (Geisinger, 2003). Additionally, comparisons of Item-Subscale correlations in this study to those reported in the SSIS-RS manual (Gresham & Elliot, 2008) could identify items that may have been altered in the translation process, at least in the way the item contributes to the subscale score (Eremenco, Cella, & Arnold, 2005). Finally, like-named scales across the two instruments were investigated to reveal any differences in subscale scores.

To summarize, the purpose of this study was to compare the Norwegian versions of the SSIS-RS and SSRS, using the same methodology as Gresham and colleagues (2011) used for their SSIS-SSRS validation. The comparisons were expected to show support for convergent and divergent validity across like-named total scales and subscales of the two instruments. Also, we expected to find good indicators for the SSIS-RSs superiority over SSRS by comparing the internal consistency estimates.

In addition, linguistic equivalence was investigated by examination of Item-Total and Item-Subscale correlations to check the Norwegian versions of SSIS-RS and SSRS for poorly performing items which could indicate translation error. Additionally, comparative analyses of the Norwegian and US items across all forms were performed to identify possible

discrepancies not necessarily identified by low Item-correlations.

We also expected to find support for conceptual equivalence based on construct validity estimates, and investigate metric equivalence of the SSIS-RS by comparing alpha coefficients found in the present study to those reported by Gresham et al., (2011). Additionally, Item-Subscale correlations found in the present study were compared to those presented in the SSIS-RS manual (Gresham & Elliot, 2008).

Finally, possible differences in subscale scores between the two instruments were investigated. As the like-named subscales of SSIS-RS and SSRS measure identical domains, similar scores between the two instruments are expected. Hence, any discrepancies would be informative of possible errors, cultural differences, or other measurement issues.

Method

Participants

Participants consisted of 599 pupils, parents and teachers from five elementary and secondary schools in two towns in northern Norway, including Tromsø and Hammerfest (Table 2). 26 elementary and secondary schools were contacted, but only five chose to participate. Ages ranged from 8-16 years (3rd to 10th grade in the Norwegian school system).

--Table 2--

This was a multi-rater survey in which three types of informants answered on behalf of one child, and the sample consisted of students, parents, and teachers completing the questionnaires. Parent and students had a response rate of 31 %. Teachers were instructed to only fill out forms that were returned back to school, and this resulted in a response rate of 86 %.

Inconsistency between reported *N* across raters presented in Table 2 is due to incomplete questionnaires and difficulties recruiting teachers for participation, especially teachers from elementary schools. The maximum level of ratings for teachers was six students

(median = 3).

Instruments

In this particular study the Norwegian versions of the SSIS-RS and SSRS were used. In addition, elementary and secondary forms were used for students from third grade to tenth grade in the Norwegian school system (age 8 to 16).

Social Skills Rating System (SSRS). SSRS (Norwegian versions) is rated on a four point scale (0 = Never, 1 = Sometimes, 2 = Often, 3 = Almost always) when assessing social skills and problem behavior on the Student, Parent and Teacher forms.

Social Skills Improvement System-Rating Scales (SSIS-RS). SSIS-RS has been translated from English to Norwegian by the means of the back-translation procedure, including forward translation and back-translation by bilinguals (Strømgren, 2012). The translation was approved by the publisher of the instrument. SSIS-RS is rated on a four point scale (0 = Never, 1 = Sometimes, 2 = Often, 3 = Almost always) when assessing social skills and problem behavior on the Parent and Teacher forms, and a four point scale (Not true, Little true, A lot true, Very true) on the Student forms, indicating how true the behaviors are for the student.

Procedure

Participants were recruited by contacting principals' in different schools in Tromsø and Hammerfest with information about the project. All schools were first contacted by e-mail, and subsequently contacted by telephone. Participation was voluntary and no extra benefits were offered**.**

In line with Gresham et al., (2011), the *Social Skills Rating System* (*SSRS*; Gresham & Elliot, 1990) and the *Social Skills Improvement System – Rating Scales* (*SSIS-RS*; Gresham & Elliot, 2008) forms were administered in counterbalanced order to cancel out any order effect.

The teacher in each class selected the 15 first students on the class list in alphabetical

order beginning with the letter A. Parents consented on behalf of their children. Students and parents filled out the forms at home and returned it back to school. The teachers were instructed to only fill out forms for student who returned the envelope after participating. Each teacher participating then filled out teacher forms for a maximum of six students. The Committee for Medical and Health Research Ethics (REK) approved the project and confidentiality of the participants was secured during the project.

Analysis

Analyses computed by *t*-tests (SSRS = 1, SSIS-RS = 2) was employed to investigate any order effect in the administration of the instruments. Little's Missing Completely at Random Test (Little, 1988) was computed to investigate if any items had significant more missing values than others.

Analysis including Pearson *r* correlations and comparison of internal consistency estimates between SSIS-RS and SSRS were done in accordance to Gresham et al., 2011. However, Gresham and colleagues also reported adjusted *r*s to correct for restriction of range, based on the variability correction of Cohen (Cohen, Cohen, West, & Aiken, 2003, p. 58). This was not possible in the present study due to the lack of norm data.

Pearson correlations *r* was computed between total scale scores and subscale scores found on SSIS-RS and SSRS for all raters to show evidence of convergent and divergent validity. Reliability estimates in the form of Cronbach's alpha were computed for all scales across all raters on SSIS-RS and SSRS, and like-named scales and subscales were compared. Comparisons were computed by *z*-test. These procedures used for comparisons required the use of Fisher's Z*r* transformation formula. Although not a common method for comparing alpha coefficients, these procedures were selected in line with Gresham et al., (2011) and performed to ensure comparability to the Gresham et al., (2011) results. Feldt (1980) recommends a test statistic using a t-distribution for comparing alpha coefficients of two tests

administered in one sample, and may be more appropriate (Feldt, Woodruff, & Salih, 1987). In the results section we present and compare both methods of analysis.

In addition, alpha coefficients of SSIS-RS found in the current study were compared to SSIS-RS alpha coefficients reported by Gresham et al., (2011). To compare Cronbach's alphas from two independent samples we used a standard procedure (Feldt et al., 1987). This statistical test uses an F-distribution. However, a bit different procedure should be used if the sample size is low or the scale consists of few items (Feldt & Seonghoon, 2006). It is considered appropriate to use this test when ($N*k$) is less than 1000 (N = sample size, k = number of items in the scale), and it controls for Type 1 error even if the values of N and k are relatively small (Feldt & Seonghoon, 2006).

Corrected Item-Total correlations were computed for all total scales and subscales across all forms of SSIS-RS and SSRS, and we used correlations below .20 to identify items not correlating sufficiently within the scales (Everitt, 2002; Field, 2005). Additionally, repeated measures ANOVA were employed to investigate any differences in scores in the social skills subscales on the Norwegian versions of SSIS-RS and SSRS.

Results

For practical reasons, tables showing validity evidence based on Pearson r correlations between Total scales and Subscales are presented here. Preliminary analysis of order effect of the instruments showed no significant differences. Also, Little's Missing completely at random test was not significant, supporting that missing data do not depend of the variables in the dataset. All total scales and subscales scores across raters were normally distributed. This indicates that the use of Persons r correlations is appropriate.

Convergent and Divergent Validity

Convergent relationships are evidenced by positive correlations between common

scales across SSIS-RS and SSRS. Also, divergent relationships are evidenced by negative correlations between social skills and problem behavior scales on the two instruments.

Teacher-elementary form. As shown in Table 3, correlations between social skills Total and Subscale scores from Teacher ratings on SSIS-RS and SSRS ($N = 26$) showed weak to very strong relations, with the strongest relations between like-named scales. Also, correlations between social skills Total and Subscale scores from SSIS-RS Teacher ratings, and problem behaviors Total and Subscale scores from SSRS Teacher ratings showed weak to strong negative relations.

However, the weak correlations were produced by the Assertion subscales on SSIS-RS, which did not correlate significantly with any of the social skills scales on SSRS, showing both weak negative and weak positive relations. In addition, the scale did not correlate negatively with the problem behavior Total scale and Subscales.

Also, as shown in Table 3, correlations between problem behaviors Total and Subscale scores from Teacher ratings on SSIS-RS and SSRS ($N = 26$) showed moderate to very strong relations, with very strong relations between like-named scales. Also, correlations between problem behaviors Total and Subscale scores from SSIS-RS and social skills Total and Subscale scores from SSRS Teacher ratings showed moderate to very strong negative relations.

--Table 3--

Teacher-secondary form. As shown in Table 4, correlations between social skills Total and Subscale scores from Teacher ratings on SSIS-RS and SSRS ($N = 155$) showed moderate to very strong relations, with strong to very strong relations between like-named scales. Also, correlations between social skills Total and Subscales scores from SSIS-RS Teacher ratings and problem behavior Total and Subscales scores from SSRS Teacher ratings showed moderate to strong negative relations.

Also, as shown in Table 4, correlations between problem behavior Total and Subscale scores from Teacher ratings on SSIS-RS and SSRS ($N = 155$) showed weak to very strong relation, with the strongest relations between like-named scales. Also, correlations between problem behaviors Total and Subscale scores from SSIS-RS and social skills Total and Subscale scores from SSRS Teacher ratings showed weak to strong negative relations.

--Table 4--

Parent-elementary form. As shown in Table 5, correlations between social skills Total and Subscale scores from Parent ratings on SSIS-RS and SSRS ($N = 53$) showed weak to strong relations with moderate to strong relations between like-named scales. Also, correlations between social skills Total and Subscale scores from SSIS-RS Parent ratings and problem behaviors Total and Subscale scores from SSRS Parent ratings showed weak to strong negative relations.

Also, as shown in Table 5, correlations between problem behaviors Total and Subscale scores from Parent ratings on SSIS-RS and SSRS ($N = 53$) showed moderate to strong relations with strong relations relations between like-named scales. Also, correlations between problem behaviors Total and Subscale scores from SSIS-RS Parent ratings and social Skills Total and Subscale scores from SSRS Parent ratings showed a moderate to strong negative relations.

--Table 5--

Parent-secondary form. As shown in Table 6, correlations between social skills Total and Subscale scores from Parent ratings on SSIS-RS and SSRS ($N = 156$) showed weak to strong relations with moderate to strong relationships between like-named scales. Also, correlations between social skills Total and Subscales scores from SSIS-RS Parent ratings and problem behaviors Total scale and Subscale scores from SSRS Parent ratings showed weak to strong negative relations.

Also, as shown in Table 6, correlations between problem behaviors Total and Subscale scores from Parent ratings on SSIS-RS and SSRS ($N = 156$) showed moderate relations with moderate relations between like-named scales. Also, correlations between problem behaviors Total and Subscale scores from SSIS-RS Parent ratings and social skills Total and Subscale scores from SSRS Parent ratings showed a weak to strong negative relations.

--Table 6--

Student-elementary form. As shown in Table 7, correlations between social skills Total and Subscale scores from Student ratings on SSIS-RS and SSRS ($N = 53$) showed moderate to strong relations with moderate to strong relations between like-named scales.

Also, as shown in Table 7, correlations between problem behaviors Total and Subscale scores from SSIS-RS Student ratings and social skills Total and Subscale scores from SSRS Student ratings showed a weak to strong negative relations.

--Table-7--

Student-secondary form. As shown in Table 8, correlations between social skills Total and Subscale scores from Student ratings on SSIS-RS and SSRS ($N = 156$) showed weak to very strong relations with strong to very strong relations between like-named scales.

Also, as shown in Table 8, correlations between problem behaviors Total and Subscale scores from SSIS-RS Student ratings and social skills Total and Subscale scores from SSRS Student ratings showed a weak to strong negative relations. Student form SSRS do not have problem behavior ratings, therefore, divergent validity investigation was not possible in the same fashion for SSIS-RS Social skills scales, only possible for problem behavior scales. Also, for the same reason, convergent validity investigations not possible for problem behavior scales on SSIS-RS student form.

--Table 8--

Internal Consistency Estimates

SSIS-RS and SSRS. Cronbach's alpha coefficients of reliability for like-named scales on SSIS-RS and SSRS were computed and the estimates were compared by z-test across the two instruments.

Teacher elementary scales ranged from $\alpha = .76$ to $\alpha = .97$. SSIS-RS alpha estimates showed no significant difference from SSRS estimates. Teacher secondary scales ranged from $\alpha = .80$ to $\alpha = .96$. Three significant differences ($p < .05$) were showed, indicating significant higher alpha coefficient for the SSRS subscales Assertion and Cooperation. Likewise, the total problem scale on SSIS-RS showed significant higher alpha coefficient than like-named scale on SSRS.

Parent elementary scales ranged from $\alpha = .64$ to $\alpha = .94$; however, only the Internalizing subscale on SSRS showed alpha coefficient below .72. Comparison showed no significant higher alpha coefficient ($p < .05$) for any like-named subscales on the two forms. Parent secondary scales ranged from $\alpha = .69$ to $\alpha = 93$; however, only the Assertion subscale on SSIS-RS showed alpha coefficient below .73. Comparison showed significant higher alpha estimates for SSRS subscales Assertion and Self-control. Likewise, the SSIS-RS problem behavior scales Externalizing, Internalizing, and Total scale showed significant higher alpha coefficients compared to like-named scales on SSRS.

Student elementary scales (only social skills scales on SSRS student form) ranged from .64 to .96, however only the subscales Self-Control and Assertion on SSRS showed alpha coefficients below .73. Two SSIS-RS scales showed significantly higher alpha estimates compared to like-named SSRS scales, including Self-control and Responsibility. No significant difference (($p < .05$) was shown for the subscales Empathy and Cooperation Student secondary scales ranged from .77 to 94. No significant difference ($p < .05$) was

shown between alpha coefficients found on the Student secondary forms.

Comparison of like-named SSIS-RS and SSRS scales were performed by *z*-test in line with the procedure chosen by Gresham et al., (2011). A total of 46 comparisons revealed 11 significantly different alpha estimates, included seven SSIS-RS scales showing higher estimates than like-named SSRS scales. Altogether, including non-significant comparisons of alpha estimates across all raters and ages, the result showed higher or equal alpha estimates for 16 of 28 SSIS-RS social skills scales and 13 of 14 SSIS-RS problem behavior scales.

By the use of the test statistic recommended by Feldt (1980) we calculated the *t*-value and significance level. This method yielded the same result as the *z*-test in terms of direction in the two-tailed analysis, but showed 28 instead of 11 significant differences; including 19 SSIS-RS scales showing significantly higher estimates then SSRS.

SSIS-RS New scales. Cronbach's alpha coefficients for all new scales and old scales now appearing on all forms on SSIS-RS compared to SSRS, showed estimates across all forms ranging from .31 to .92. Except for a low alpha coefficient of .31 produced for the Bullying scale on the Student elementary form, the only alphas below .70 was the Communication (α = .68) and Bullying (α = .68) subscales found on the Student secondary form.

SSIS-RS Norwegian and US-version. Cronbach's alpha coefficients of reliability for SSIS-RS scales presented in Gresham et al., (2011) were compared to alpha coefficients produced in the present study to investigate metric equivalence. Comparisons by *F*-test of 46 total scales and subscales from all forms showed a significant difference (p < .05) between 14 alpha coefficients.

Corrected Item-Total correlations. Corrected Item-Total correlation ranged from -.25 to .92 for Subscales and Total scales across all forms on SSIS-RS and SSRS. Investigation of Item-Subscale Correlations identified 8 items on SSIS-RS and 10 items on SSRS with

estimates below .20, and Item-Total scale identified 14 items on SSIS-RS and 11 items on

SSRS below this cut off. Nearly all the low correlations (< .20) were found on the elementary

forms, including 19 of 22 items on SSIS-RS and 20 of 21 on SSRS. Furthermore, two items

distinguished themselves with low Item-Total correlations found across both age-levels,

including item 1 and 41 found on the SSIS-RS Parent forms in the Assertion and Self-Control

scale, respectively. The US item 1 sounded, "Expresses feelings when wronged" as opposed

to the Norwegian translation "Utrykker følelser ved urett". Further, the US item 41 was

"Tolerates peers when they are annoying", as opposed to the Norwegian translation "Tolererer

andre som erter og plager".

We also compared Item-Subscale correlations on the Secondary forms found in this

study with those reported in the SSIS-RS manual (Gresham & Elliot, 2008) to search for

items that may have been altered in the translation process, but not necessarily have Item-

Total correlations below .20 (Eremenco et al., 2005). Item-Subscale correlations were highly

comparable. However, two items distinguished themselves; item 36 on the Parent self-control

scale "Makes a compromise during a conflict" vs. "Kompromisser i konflikter" and item 71

on the Student internalizing scale "I feel nervous with my classmates" vs. "Jeg er redd for

mine klassekamerater". Comparison was computed by transformation of each Corrected Item-

Subscale correlation via the Fisher Zr transformation formula. These showed large differences by a

much lower estimate produced for the Norwegian items ($Z = 4.45$, $p < .001$ and $Z = 6.31$, $p <$

.001, respectively).

Levels of subscale scores between SSIS-RS and SSRS. Levels of like-named social

skills subscale scores between SSIS-RS and SSRS were subjected to repeated measures

ANOVA for all raters. Almost all results evidenced a high degree of similarity between the

two instruments. One exception was found on the Parent secondary forms, where the ANOVA

indicated a significant effect of the two instruments over the four subscales, $F\ (3,\ 468) =$

95.39, $p < .0001$. The difference was traced to lower scores on the SSRS Cooperation scale. Specifically, three specific items asking about chores at home in the SSRS had significantly lower estimates than the other items comprising the scale.

Discussion

The purpose of this study was to compare the *Social Skills Improvement System-Rating Scales* (*SSIS-RS*: Gresham & Elliott, 2008) with the *Social Skills Rating System* (*SSRS*; Gresham & Elliott, 1990). Comparison of the two instruments focused on validity and reliability estimates for all three forms (teacher, parent, and student) in both elementary and secondary school.

Validity. Pearson *r* correlations across Total scale and Subscales of the two instruments were as predicted. The results indicated moderate to very high positive relations between social skills scales on SSIS-RS and SSRS and between problem behaviors scales on SSIS-RS and SSRS; therefore, convergent validity was indicated. In addition, Pearson *r* correlations across social skills scales and problem behavior scales showed moderate to very strong negative relations, indicating divergent validity. Furthermore, these results support the construct validity of SSIS-RS.

However, an exception to this result was the SSIS-RS Assertion scale on the Teacher elementary form, which did not correlate positively with the social skills scales, or negatively with the problem behavior scales on the SSRS Teacher elementary form. A likely reason for this result was the small sample size used for analysis of the Teacher elementary form ($N = 26$). Additionally, the Assertion scale on some of the other forms also showed weaker positive relationships with SSRS social skills scales and weaker negative relationships with SSRS problem behavior scales compared with the other social skills scales on SSIS-RS. This overall lower correlation coefficients produced by the Assertion scale was also reported by Gresham et al., (2011). A possible explanation for this lower correlation is the ambiguity which may

rise when judging a child's assertiveness: On one hand, assertive behavior can be judged as a positive action (e.g. express feelings when wronged; stands up for others when treated unfairly). On the other hand, these same assertive behaviors could also be interpreted as problematic behavior in some situations or circumstances (Argyle, Furnham, & Graham, 1983).

Internal Consistency. Based on the test review form developed by the European Federation of Psychologists Association (EFPA), a criteria for use in assessment is an adequate internal consistency (Cronbach`s α) of > .70 (European Federation of Psychologist Associations, 2008). Therefore, in our study the overall internal consistency was acceptable for all SSIS-RS like-named total scales and subscales on all forms, indicating that the items comprising each scale are measuring the same construct. The only like-named scale below α = .70 on SSIS-RS was the Assertion scale on the Parent secondary form (.69). However, lower estimates for the Assertion scale compared to other social skills subscales is consistent with the results reported by others (Gresham & Elliot, 2008; Gresham et al., 2011). Further, these alpha estimates might reflect the already mentioned ambiguity when rating a child's assertive behavior.

Comparison between SSIS-RS and SSRS. Comparison of alpha coefficients between all like-named scales on SSIS-RS and SSRS showed high consistency between the instruments. Only four like-named subscales showed significantly higher alpha coefficient in favor of SSRS. However, two of these scales were the Cooperation and the Assertion scale on Teacher secondary form and both these scales still showed high alphas. This result is also similar to Gresham et al., 2011, reporting significantly lower alpha estimate for SSIS-RS Cooperation scale and a non-significant difference between the Assertion scales.

Additionally, seven SSIS-RS scales showed higher alpha coefficients compared to like-named scales on SSRS. This result is also in line with Gresham et al., (2011), but in

comparison, they reported significant differences between 35 of 46 pairs of alpha coefficients in favor of SSIS-RS, and only two SSRS scales showed significantly higher estimates. This may indicate an overall higher internal consistency for the US SSIS-RS.

Further, the SSIS-RS social skill subscales are shorter then SSRS social skill subscales and mostly produce similar or larger alpha coefficients; therefore, supporting an advantage of SSIS-RS over SSRS. However, problem behavior scales found on SSIS-RS have more items comprising each scale compared to like-named scales on SSRS. Therefore, were the results showed similar or lower alpha coefficients the opposite could be indicated.

We also note that comparison by the *t*-test recommended by Feldt (1980) showed additional significant differences between the instruments in our study. This implies that the result reported by Gresham and colleagues (2011) would show stronger results in the same direction using Feldt's method; therefore, continuing to show better Cronbach`s alpha estimates for SSIS-RS.

New subscales. All new subscales on SSIS-RS, in addition to already existing subscales now present on all forms (e.g. Responsibility and Empathy subscales) showed mostly acceptable alpha coefficients. However, the low alpha coefficients on the Bullying scale on the Student forms may indicate that the students have difficulties reporting their own bullying behavior consistently, especially elementary school students ($\alpha = .31$). Further, such low estimates were not shown for the US or the Spanish version reported by Gresham & Elliot (2008). We could not identify any obvious problems with the translation, including investigation of Corrected Item-Total correlations produced for the Bullying scale, but the scale comprises of only five items and therefore may be especially vulnerable to atypical data.

Cross-cultural equivalence. *Linguistic equivalence.* Corrected Item-Total scale and Item-Subscale correlations were generally moderate to strong for all items on SSIS-RS and SSRS for the secondary forms, and nearly all low correlations (< .20) were shown for items

on the elementary forms. A probable reason for this is the low sample sizes used analyzing the elementary forms. Additionally, almost all items showing low correlations on the elementary forms are also present on the same rater's secondary forms, but did not show the same result. This indicates that the items function well in their respective scales on the secondary forms. Also, a possible reason for more inconsistencies found on the elementary forms is that it might be harder to report younger children's behavior consistently due to more variability in development and behavioral expression (Merrell, 2003).

Two items on SSIS-RS (item 1 and item 41) which revealed low Item-Total correlations on both age levels had differences in meaning across the two languages. The US item 1 was more objective, "Tolerates peers when they are annoying", and could therefore be reported from a general perspective, whereas the Norwegian translation sounded more subjective "Tolererer andre som erter og plager". Contrary, item 41 ("Expresses feelings when wronged" as opposed to the Norwegian translation "Utrykker følelser ved urett") could be interpreted as more specific and personally compared to the Norwegian item, which may also include general injustice. This implies that the difference in meaning between the two languages could be the reason for the different results. Furthermore, these low correlations produced by item 1 and 41 were not in line with Gresham and Elliot (2008) reporting high Item-Subscale correlations for all items on the US and Spanish Parent forms.

Additional translation errors must be noted. Item 4 on the Norwegian SSRS Parent elementary form were replaced; originally sounding "The child will join group activities without being told to" was changed to "Gir rimelig uttrykk for skuffelse når han/hun ikke lykkes". In addition, two items on SSIS-RS Student elementary form and two items on SSRS Student secondary form were practically identical after translation. Including items 1 and 45 on SSIS-RS "Jeg spør om hjelp når jeg trenger det" originally being "I ask for information when I need it" and "I ask for help when I need it". Also, items 2 and 24 on SSRS are

practically identical "Jeg roser andre når det har gjort noe bra" and "Jeg roser andre hvis jeg synes de gjør noe bra" originally sounding "Say nice things to others" and "Tell others when they have done well".

Generally, there seems to be support for linguistic equivalence for all SSIS-RS secondary forms. However, the discrepancies mentioned should be considered before further use of the instrument in Norway. Because of low sample size used in the investigation of the elementary forms we found it difficult interpret evidence of linguistic equivalence. Nevertheless, considering the high similarity of the elementary and secondary forms for each rater, this supports the possibility that low Corrected Item-Total correlations shown for the elementary forms are not due to translation errors.

Conceptual equivalence. Support for conceptual equivalence was established with evidence of convergent and divergent validity, indicating that the translated construct are related to the construct it is supposed to; therefore, supporting that the meaning is cared for in the translation process. Conceptual equivalence is important to establish because metric equivalence presumes conceptual equivalence (Geisinger, 2003).

Metric equivalence. 32 of 46 comparisons of alpha coefficients found in the present study to those reported by Gresham et al., (2011) showed no significant difference, indicating metric equivalence. Additionally, Item-Subscale correlations on the secondary forms were compared to US Item-Subscale correlations reported in the manual (Gresham & Elliot, 2008) to uncover items, which might have been altered in the translation process. The comparisons showed high similarities, with the exceptions of item 36 and 71. Further examination of these items presented some possible translation inaccuracies. Item 36 was "Kompromisser i konflikter" in Norwegian, and might sound incomplete or create some uncertainty compared to the original item sounding; "Makes a compromise during a conflict". Item 71 originally sounds "I feel nervous with my classmates" as opposed to the Norwegian translation "Jeg er

redd for klassekameratene mine" which in our opinion clearly can be interpreted differently. These items should be considered altered in future studies.

We did not perform this comparison with the elementary forms because of the many low Corrected Item-Total correlations, as reported earlier. This implies that metric equivalence has been partly supported. However, 14 subscales showed significantly different alpha coefficients, indicating additional investigation in a different sample.

Differences in subscale scores between SSIS-RS and SSRS. Repeated measures ANOVA revealed a difference in subscale scores on the Cooperation scale on the Parent secondary form. This was at least in part due to the fact that items on SSRS Parent Cooperation scale are more concrete as opposed to more generally formulated items on SSIS-RS Parent Cooperation scale. Especially three items about chores at home yielded lower scores. This implies that investigation of instrument items can be important when assessing inter-rater reliability, which often are reported low because of rater's different perspective and information (Achenbach et al., 1987; Renk & Phares, 2004).

Identical items on the Parent elementary form did not demonstrate similar differences, and the subscale scores found on the Parent elementary forms were equally high. However, these behaviors can be more noticeable among older children. This indicates the importance of a clear conceptualization of the items comprising a subscale, in addition to the impact variables like age may have.

Improvement made on the basis of earlier critics indicates that the SSIS-RS developers have addressed many documented technical concerns (Frey et al., 2011). According to Gresham et al., (2011) the new subscales in SSIS-RS were developed in line with the same procedures and professionals which have outlined the rest of the SSIS-RS and therefore should expect the same psychometric properties. Furthermore, SSIS-RS is built on the theoretical basis of SSRS which have much empirical support for its validity and reliability.

The validity assumptions in this study were supported by showing convergent and divergent relations indicated by positive and negative correlations where it was expected. Further, Cronbach`s alpha coefficients were generally acceptable supporting the internal consistency. However, estimates for the new Bullying scale on the Student forms indicated a need for further investigation. In addition, although we have proposed some considerations regarding the translation of some items, the overall evidence for cross-cultural equivalence is promising.

Limitations

Sample size. Low sample size is a possible problem for the validity investigation of the Assertion scale on the Teacher elementary form. However, we believe that the sample size may not be crucial for validity estimates since the remaining correlations between the two instruments are as expected. Also, as mentioned earlier, the Assertion scale may be more ambiguous to rate; therefore, a larger sample size would probably be beneficial for this investigation. Further, this notion was supported by analysis of the Assertion scale in larger samples, including the investigation of the secondary forms and results reported by others (Gresham et al., 2011; Gresham & Elliott, 2008). In addition, sample size could also be a reason for lower Corrected Item-Total correlations produced across all elementary forms, compared to secondary forms. Therefore, based on the investigations of Corrected Item-Total correlations on the secondary forms, we conclude that this particular investigation on the elementary forms should be performed with a larger sample to render results that are more conclusive.

Representativeness. At all levels of agreeing to participate, self-selection was involved. Whether a given participant in fact took part depended on the willingness to participate from three persons: First the student must bring the questionnaire home, then, both parents and student must choose to participate, and finally the forms must be returned to

school. Lastly, the teacher decides on whether to participate or not. Therefore, in the present

study, the procedure we chose in combination with SSIS-RS and SSRS being multi-informant

instruments may have caused response rate to be artificially low.

Overall, it can be argued that the low response rate and the limited control over the

respondents might have caused the representativeness to be threatened. Additionally, the

participants in this study might have been in the upper segment of the social competence

level, and lower segment of the reported problem behaviors. This may indicate that the

generality of our conclusions might be limited, and generalizations should be done with

caution. However, as the purpose of our validation was to investigate the comparability

between the two instruments based on correlations, the level of competence or problem

behavior of the participants, or the variance in the sample, may not affect the result.

Future Research

In a future perspective, the Norwegian version of SSIS-RS should continue to be

validated against other well-established instruments used in Norway for assessment of social

behavior and school functioning, e.g. the *Strength and Difficulties Questionnaire* (*SDQ*;

Goodman, 1997), and the *Child Behavior Checklist* (*CBCL); Teacher Report Form* (*TRF);

Youth Self-Report* (*YSR)*, which are all parts of *The Achenbach System of Empirically Based

Assessment* (*ASEBA*; Achenbach & Rescorla, 2001). All these instruments focus on problem

behavior, pro-social behavior and social competence. Furthermore, it is also important to

validate the new subscale on the SSIS-RS, e.g. the autism subscale against the *Autism

Spectrum Rating Scales* (*ASRS;* Goldstein & Naglieri, 2009), and the Bullying scale against

OLWEUS (Olweus, 1996). Additionally, social validity is concerned with the social value of

changes in behavior being measured (Wolf, 1978); therefore, the important ratings on the

SSIS-RS social skills forms could be investigated in a Norwegian sample to investigate social

validity. Likewise, a validation against other methods of social behavior assessment (e.g. direct observations) would be beneficial.

For several reasons, a future aim would be creating a norm standardization sample. First, US norms may not be representative in a Norwegian context based on cultural differences (Rescola et al., 2007). Second, from a large representative sample a factor analysis would empirically investigate the structure of the scales and subscales in a more conclusive fashion than possible in the present study. Also, comparisons of factor loadings from a US study could further indicate metric equivalence. A test-retest is also of interest in order to examine the SSIS-RSs stability over time. Evidently, there is also a need to validate a Sami version of the SSIS-RS based on cultural differences (Javo, Rønning, Handegård, & Rudmin, 2009).

Conclusion

Based on the empirical evidence from this study the SSIS-RS appears to meet satisfactory construct validity measures, and it is likely to claim the instrument assessing the same constructs as the SSRS. Being built on the same foundation as its predecessor, and by recognizing and addressing the concerns considering the SSRS, the SSIS-RS is likely to produce valid and reliable scores for social behaviors.

References

Achenbach, T., McConaughy, S., & Howell, C. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232.

Achenbach, T., & Rescorla, L. (2001). *Manual of the ASEBA school-age forms and profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, and Families.

Albertus, C., Birkinbine, J., Lyon, M. A., & Naibi, J. (1996). A validity study of the social skills rating system-teacher version with disabled and nondisabled preschool children. *Perceptual and Motor Skills, 83*, 307-316.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Argyle, M., Furnham, A., & Graham, J.A. (1983). *Social situations.* Cambridge, England: Press Syndicate

Bjørnebekk, G., & Howard, R. (2012). Validation of a motivation-based typology of angry aggression among antisocial youths in Norway. *Behavioral Sciences and the Law, 30,* 167-180.

Bramlett, R. K., Smith, B. L., & Edmond, J. (1994). A comparison of nonreferred, learning-disabled, and mildly retarded students utilizing the social skills rating system. *Psychology in the Schools, 31*, 13-19.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the Multitrait-Multimethod Matrix. *Psychological Bulletin*, *56*, 81-105.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple*

regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Erlbaum.

Cowen, E., Pederson, A., Babigian, H., Izzo, I., & Trost, M. (1973). Long term follow-up of early detected of vulnerable children. *Journal of Consulting and Clinical Psychology, 41*, 1400-1416.

Crowe, L. M., Beauchamp, M. H., Catroppa, C., & Anderson, V. (2011). Social function assessment tools for children and adolescents: A systematic review from 1988 to 2010. *Clinical Psychology Review, 31*, 767-785.

Demaray, M., Ruffalo, S., Carlson, J., Busse, R., Olson, A., McManus, S., & Leventhal, A. (1995). Social skills assessment: A comparative evaluation of six published rating scales. *School Psychology Review, 24*, 648-671.

Dodge, K. A., & Pettit, G. S. (2003). Biopsychosocial model of the development of chronic conduct problems in adolecence. *Developmental Psychology, 39*, 349-371.

Elliot, S. N., & Gresham, F. M. (1987). Children's social skills: Assessment and classification practices. *Journal of Counseling and Development, 66*, 96-99.

Eliassen, S. (2012). *Aggression Replacement Training: Deltakernes vurdering av motivasjonelle faktorer* (Unpublished master`s thesis). Department of Psychology, University of Tromsø, Tromsø.

Eremenco, S., Cella, D., & Arnold, B. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Proffessions, 28*, 212-232.

European Federation of Psychologist Associations (2008). *Test review forms and notes for reviewers.* Retrieved from European Federation of Psychologist Associations website: http://www.efpa.eu/profesionale-development

European Federation of Psychologist Associations (2009). *Reviewing translated and adapted tests.* Retrieved from European Federation of Psychologists Associations website: http://www.efpa.eu/profesionale-development

Everitt, B. S. (2002). *The Cambridge Dictionary of Statistics* (2nd ed.). Cambridge: CUP.

Feldt, L. S. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. *Psychometrika, 45*, 99-105.

Feldt, L. S., & Seonghoon, K. (2006). Testing the difference between two alpha coefficients with small samples of subjects and raters. *Educational and Psychological Measurement, 66*, 589-600.

Feldt, L. S., Woodruff, D. J., & Salih, F. A. (1987). Statistical inference for coefficient alpha. *Applied Psychological Measurement, 11*, 93-103.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage.

Flanagan, D., Alfonso, V., Primavera, L., Povall, L., & Higgins, D. (1996). Convergent validity of the BASC and SSRS: Implications for social skills assessment. *Psychology in the Schools, 33*, 13-23.

Frey, J. R., Elliott, S. N., & Gresham, F. M. (2011). Preschoolers' social skills: Advances in assessment for intervention using social behavior ratings. *School of Mental Health, 3*, 179-190.

Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In Irving B. Weiner (Eds.), *Handbook of psychology* (pp. 95-118). New Jersey: John Wiley & Sons, Inc.

Goldstein, S., & Naglieri, J. A. (2009). *Autism Spectrum Rating Scales*. Toronto, Canada: Multi-Health Systems.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586.

Gresham, F. M. (1992). Social skills and learning disabilites: Causial, comcomitant or correlational. *School Psychology Review, 21*, 348-360.

Gresham, F. M., Cook, C. R., Crews, S. D., & Kern, L. (2004). Social skills training for children and youth with emotional and behavior disorders: Validity considerations and future directions. *Behavioral Disorders, 30*, 32-46.

Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating System.* Circle Pines, MN: American Guidance service.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System Rating Scales manual*. Minneapolis, MN: NCS Pearson.

Gresham, F. M., Elliot, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010a). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skill Improvement System- Rating Scales. *Psychological Assessment, 22*, 157-166.

Gresham, F. M., Elliot, S. N., Vance, M. J., & Cook, C. R. (2011). Comparability of the Social Skill Rating System to the Social Skills Improvement System: Content and psychometric comparisons across elementary and secondary age levels. *School Psychology Quarterly, 26*, 27-44.

Gresham, F. M., & Reschly, D. J. (1987). Dimensions of social competence: Method factors in the assessment of adaptive behavior, social skills, and peer acceptance. *Journal of School Psychology, 25*, 367-381.

Gudykunst, W. B (2002). Issues in cross-cultural communication research. In: W.B. Gudykunst, & B. Mody (Eds.), *Handbook of International and Intercultural Communication* (2nd ed., pp. 165-179). Thousand Oaks, CA: Sage.

Gundersen, K., & Svartdal, F. (2006). Aggression Replacement Training in Norway: Outcome evaluation of 11 Norwegian student projects. *Scandinavian Journal of Educational*

Research, 50, 63-81.

Humphrey, N., Kalambouka, A., Wigelsworth, M., Lendrum, A., Deighton, J., & Wolpert, M.

(2011). Measures of social and emotional skills for children and young people: A

systematic review. *Educational and Psychological Measurement, 71*, 617-637.

Javo, C., Rønning, J. A., Handegård, B. H., & Rudmin, F. W. (2009). Social competence and

emotional/behavioral problems in a birth cohort of Sami and Norwegian

preadolescents in Arctic Norway as reported by mothers and teachers. *Nordic Journal*

of Psychiatry, 63, 178-187.

Jurado, M., Cumba-Avilés, E., Collazo, L. C., & Matos, M. (2006). Reliability and validity of

a Spanish version of the Social Skills Rating System-teacher form. *Journal of*

Psychoeducational Assessment, 24, 195-209.

Kankaras, M., & Morse, G. (2010). Researching measurement equivalence in cross-cultural

studies. *Psihologija, 43*, 121-136.

Kohn, M., & Clausen, J. (1955). Social isolation and schizofrenia. *American Sociological*

Review, 20, 265-273.

Kupersmidt, J. B., Coie, J. D., & Dodge, K. A. (1990). The role of poor peer relationships in

the development of disorder. In S. R Asher & J. D. Coie (Eds*.), Peer rejection in*

childhood (pp. 274-308). New York: Cambridge University Press.

Langeveld, J. H., Gundersen, K., & Svartdal, F. (2011). Social competence as a mediating

factor in reduction of behavioral problems. *Scandinavian Journal of Educational*

Research, 56, 1-19.

Levinson, D. (1998). *Ethnic groups worldwide: A ready reference handbook*. Phoenix,

Arizona: The Oryx Press.

Little, R. J. A. (1988). A test of missing completely at random for multivariate data with

missing values. *Journal of the American Statistical Association, 83*, 1198-1202.

Maag, J. W. (2005). Social skills training for students with emotional behavioral disorders

and learning disabilites: Problems, conclusions, and suggestions. *Exceptionality, 13*,

155-172.

Manz, P. H., Fantuzzo, J., & McDermott, P. (1999). The parent version of preeschool Social

Skills Rating Scale: An analysis of its use with low-income, ethnic minority children.

School Psychology Review, 28, 493-504.

Merrell, K. W. (2003). *Behavioral, social, and emotional assessment of children and

adolescents*. Mahwah, NJ: Lawrence Erlbaum Associates.

Merrell, K. W., & Popinga, M. R. (1994). The alliance of adaptive behavior and social

competence: An examination of relationships between the skills of independent

behavior and the social skills rating system. *Research in Developmental Disabilities,

1*, 39-47.

Newcomb, A., Bukowski, W., & Pattee, L. (1993). Childrens peer relations: A meta-analytic

review of popular, rejected, neglected, controversial, and average sociometric status.

Psychological Bulletin, 113, 306-347.

Nyquist, G.J., & Rasch-Olsen, I.S. (2011). *Depresjonssymptomer hos ungdom: Sosial

kompetanse-en beskyttelsesfaktor for ungdom som vokser opp i hjem med multippel

risiko?* (Unpublished master`s thesis). University of Oslo, Oslo.

Ogden, T. (2003). The validity of teacher ratings of adolecents' social skills. *Scandinavian

Journal of Educational Research, 47*, 63-76.

Ogden, T. (2011). Sosial ferdighetsopplæring for barn og ungdom. *Tidskrift for Norsk

Psykologiforening, 48*, 64-68.

Ogden, T., & Hagen, K. A. (2008). Treatment effectiveness of parent management training in

Norway: A randomized controlled trial of children with conduct problems. *Journal of

Consulting and Clinical Psychology, 4*, 607-621.

Ogden, T., & Halliday-Boykins, C. A. (2004). Multisystematic treatment of antisocial

adolescents in Norway: Replication of clinical outcomes outside of the US. *Child and*

Adolescent Mental Health, 9, 76-82.

Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire.* Mimeo. Bergen,

Research Center for Health Promotion, University of Bergen.

Parker, J. G., & Asher, S. R. (1987). Peer relations and later personal adjustment: Are low

accepted children at risk? *Psychological Bulletin, 102*, 357-389.

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and

adolescents. *Clinical Psychology Review, 24*, 239-254.

Rescorla, L., Achenbach, T., Ivanova, M. Y., Dumenci, L., Almqvist, F., Bilenberg, N., &

Verhulst, F. (2007). Behavioral and emotional problems reported by parents of

children ages 6 to 16 in 31 societies. *Journal of Emotional and Behavioral Disorders,*

15, 130-142.

Shahim, S. (2001). Reliability of the Social Skills Rating System in a group of Iranian

children. *Psychological Reports, 89*, 566-570.

Smith, B., Barkley, R., & Shapiro, C. (2007). *Attention-deficit/hyperactivity disorder.*

Assessment of childhood disorders (4th ed.). New york: Guilford Press.

Statistics Norway. (2013). Population by age, sex, marital status and citizenship. 1.

Jaunuary 2013. Retrieved from

http://www.ssb.no/en/befolkning/statistikker/folkemengde

Strømgren, B. (2012). *Norwegian translation of Social Skills Improvement System - Rating*

Scales, Unpublished manuscript, Institute for Behavior Science, Høgskolen i Oslo og

Akershus, Oslo, Norway.

Svartdal, F., & Klaussen, T. (2013). *Måling av sosial kompetanse og problematferd: En vurdering av tilgjengelige instrumenter*. Unpublished manuscript, Department of Psychology, University of Tromsø, Tromsø, Norway.

Sœstad, T., & Kyrrestad, H. (2007). *SSRS questionnaires administered over the internet: Aviable alternative for assessing children's social skills* (Unpublished master's thesis). University of Tromsø, Tromsø.

Sørlie, M-A., Amlund Hagen, K., & Ogden, T. (2008). Social competence and antisocial behavior: Continuity and distinctiveness across early adolescence. *Journal* of *Research* on *Adolescence*, *18*, 121-144.

Van der Oord, S., Van der Meulen, E. M., Prins, P. J. M., Oosterlaan, J., Buitelaar, J. K., & Emmelkamp, P. M. G. (2005). A psychometric evaluation of the social skill rating system in children with attention deficit hyperactivity disorder. *Behaviour Research and Therapy, 43*, 733-746.

Vasil'ová, K., & Baumğartner, F. (2004). *Why is social intelligence difficult to measure?* Retrieved from http://www.saske.sk/cas/archiv/4-2005/studia_vasilova-baumgartner.html

Walthall, J. C., Konold, T. R., & Pianta, R. C. (2005). Factor structure of the Social Skills Rating System across child gender and ethnicity. *Journal of Psychoeducational Assessment, 23*, 201-215.

Whiteside, S. P., McCarthy, D. M., & Miller, J. D. (2007). An examination of the factor structure of the Social Skills Rating System Parent Elementary Form. *Assessment, 14*, 246-254.

Wolf, M. M. (1978). Social validity: The case for subjective measurment or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203-214.