Title: Validation of the Strengths and Difficulties Self-Report in Norwegian Sign Language

Authors: Chris Margaret Aanondsen[1,2], Thomas Jozefiak[1], Kerstin Heiling[3] and Tormod Rimehaug[1,4]

Affiliations:

[1] Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU)

[2]Unit for deaf and hard-of-hearing children and adolescents in Central Norway, Department of Child and Adolescent Psychiatry, St. Olavs Hospital Trondheim University Hospital, Trondheim, Norway

[3]Department of Child and Adolescent Psychiatry, IKVL, Psychiatry Skåne, Lund University, Sweden

[4]Department of Child and Adolescent Psychiatry, Nord-Trøndelag Hospital Trust, Levanger, Norway

Corresponding author:

Chris Margaret Aanondsen

RKBU Midt-Norge, NTNU

Postboks 8905 MTFS

7491 Trondheim; Norway

Tel +47 73 55 15 22

Email: chris.m.aanondsen@ntnu.no

Abstract

The majority of studies on mental health in deaf and hard-of-hearing children report a higher level of mental health problems. Inconsistencies in reports of prevalence of mental health problems have been found to be related to a number of factors such as language skills, cognitive ability, heterogeneous samples as well as validity problems caused by using written measures designed for typically hearing children. This study evaluates the psychometric properties of the self-report version of the Strengths and Difficulties Questionnaire (SDQ) in Norwegian Sign Language (NSL; SDQ-NSL) and in written Norwegian (SDQ-NOR). Forty-nine DHH children completed the SDQ-NSL as well as the SDQ-NOR in randomized order while their parents completed the parent version of the SDQ-NOR and a questionnaire on hearing and language-related information. Internal consistency was examined using Dillon-Goldstein's rho, test-retest reliability using intraclass correlations, construct validity by confirmatory factor analysis (CFA) and partial least squares structural equation modeling. Internal consistency and test-retest reliability were established as acceptable to good. CFA resulted in a best fit for the proposed five-factor model for both versions, although not all fit indices reached acceptable levels. The reliability and validity of the SDQ-NSL seem promising even though the validation was based on a small sample size.

Running Head (shortened title): Validation of the SDQ-NSL and SDQ-NOR

Two reviews and a meta-analysis have reported an elevated prevalence of emotional and behavioral problems in deaf and hard-of-hearing (DHH) children and adolescents across countries, informants and measures (Fellinger, Holzinger, & Pollard, 2012; Stevenson, Kreppner, Pimperton, Worsfold, & Kennedy, 2015; Theunissen et al., 2014). For brevity, the term "children" will be used to describe both children and adolescents in this paper. The majority of studies have reported that 20−50% of DHH children suffer from mental health problems (Dammeyer, 2010b; Fellinger, Holzinger, Sattel, & Laucht, 2008; Hintermair, 2007; van Eldik, 2005; van Eldik, Treffers, Veerman, & Verhulst, 2004; van Gent, Goedhart, Hindley, & Treffers, 2007) whereas Sinkkonen (1994) reported rates comparable to those of typically hearing (TH) children based on teacher-reports. Mejstad, Heiling, and Svedin (2009) found equivalent rates of emotional and behavioral problems in DHH and TH boys based on the self-report version of the Strengths and Difficulties Questionnaire (SDQ).

Fellinger, Holzinger, Sattel, Laucht, and Goldberg (2009) found point and lifetime prevalence rates of 32.6% and 45.3%, respectively, for any psychiatric disorder in a representative Austrian DHH sample of children. Theunissen et al. (2014) concluded in their systematic review that DHH were more likely to suffer from depression, aggression, oppositional defiant disorder and conduct disorder than their TH peers. A possible cause for differences in prevalence rates found for DHH children are heterogeneous samples as well as different inclusion criteria across studies such as different degrees of hearing loss and modes of communication. Additional disabilities, communicative skills, and intellectual functioning have been shown to affect DHH children's mental health whereas the degree of hearing loss has not (Dammeyer, 2010b; Fellinger et al., 2009; Hintermair, 2006; Mejstad et al., 2009; Stevenson et al., 2017; van Gent et al., 2007). Mejstad et al. (2009) suggested that the extent of parental support provided in Sweden and Finland may ensure better mental health. This is

in accordance with Dammeyer's (2010b) study, which found no increased risk of mental health problems in DHH children with good signing or oral communication skills.

The Strengths and Difficulties Questionnaire (SDQ)

The SDQ is a brief measure to assess emotional and behavioral problems and pro-social behavior in children. It consists of 25 items that are grouped into five scales (Emotional Problems, Conduct Problems, Hyperactivity-Inattention, Peer Problems, and Pro-social Behavior). The SDQ is a multi-informant assessment and can be completed by parents of 4−17-year-olds, teachers of 4−17-year-olds and 11−17-year-old adolescents. Achenbach, McConaughy, and Howell (1987) have emphasized the importance of multi-informant assessments for capturing the unique perspectives held by each informant. The original validation demonstrated satisfactory reliability (internal consistency and test-retest reliability) and validity for all informants. The Peer Problems scale showed the lowest internal consistency (α=.41) for the self-report (R. Goodman, 2001).

The SDQ is available free of charge in over 80 languages and has been used in community and clinical samples across the world. Essau et al. (2012) compared the psychometric properties of the self-report SDQ across five European countries (the UK, Germany, Sweden, Italy, and Cyprus). They report good to satisfactory internal consistency for most subscales in most countries, with the lowest for Conduct and Peer Problems. Confirmatory factor analysis (CFA) showed that the three-factor model (internalizing and externalizing difficulties and Pro-social Behavior) demonstrated best fit in Cyprus whereas the five-factor model showed a better fit in Germany, the UK, and Sweden. The model fit indices for the five-factor model in Sweden and the UK, however, did not reach acceptable levels. A. Goodman, Lamping, and Ploubidis (2010) examined the fit of the three and five-

factor models in a large British sample and concluded that the five-factor model should be maintained for clinical samples whereas the three-factor model may be better suited to assess low-risk community samples.

The reliability and validity of the Norwegian SDQ self-, parent- and teacher-reports were found to be acceptable (Rønning, Handegaard, Sourander, & Mørch, 2004; Sanne, Torsheim, Heiervang, & Stormark, 2009; van Roy, Groholt, Heyerdahl, & Clench-Aas, 2006; van Roy, Veenstra, & Clench-Aas, 2008). Van Roy et al. (2008) demonstrated acceptable psychometrics of the self-report version for adolescents aged 11 to 19 years although the internal consistency for Conduct Problems was low for all adolescents (α = .44 to .54).

The SDQ in DHH samples

In their meta-analysis, Stevenson et al. (2015) found an elevated rate of emotional and behavioral difficulties in DHH children based on parent and teacher SDQ reports. The most pronounced risk was found for Peer Problems for informants, whereas Hyperactivity-Inattention did not show an elevated level for either of the informants. Stevenson et al. (2015) further argue that the less elevated rates in the SDQ-studies in the meta-analysis as compared to the non-SDQ-studies may reflect an actual improvement in the provision of services as a number of the non-SDQ studies were published much earlier than the SDQ-studies.

The psychometric properties of the written SDQ for DHH children have been examined in Denmark (Niclasen & Dammeyer, 2016) and Germany (Hintermair, 2007). Niclasen and Dammeyer (2016) concluded that the five-factor model could be recommended for DHH children, in a bilingual/bicultural and an oral/mainstream setting, in Denmark, with better model fit demonstrated for the teacher- than the parent-report. Hintermair (2007) found acceptable internal consistency for most subscales except for Conduct Problems (α = .51) as well as support for the five-factor model for the parent-report in Germany.

Studies have reported difficulties in reading in many DHH children (Harris, Terlektsi, & Kyle, 2017; Marschark, Convertino, et al., 2007; Marschark et al., 2009), which in turn will affect their ability to complete written forms and the validity of the results. Therefore, the SDQ has been translated to British (BSL) and Australian Sign Language (Auslan). Cornes et al. (2006) found acceptable test-retest reliability (SDQ-Auslan: .75 to .85) and internal consistency (written: .53 to .84, SDQ-Auslan: .42 to .83) for the self-report SDQ. The SDQ-Auslan, however, demonstrated higher internal consistency for all subscales except for Peer Problems (.42). Peer Problems and Conduct Problems (.55) were found to have the lowest consistency for the Auslan version; construct validity of the SDQ was not assessed. Further, Cornes et al. (2006) found no significant correlations between the written and Auslan versions for Emotional Problems (.29), Conduct Problems (.27) and Hyperactivity-Inattention (.31) subscales. Significant correlations were found for Peer Problems (.43), Pro-social Behavior (.44) and Total Difficulties (41). Significant correlations were also found for all subscales and total score between the parent- and the written self-report (.34 to . 66), whereas only Hyperactivity-Inattention (.41), Peer Problems (.35) and Total Score (.39) were significantly correlated for the SDQ-Auslan and parent-report.

Roberts et al. (2015) reported that the BSL versions of the self-, parent- and teacher-report demonstrated similar reliability and validity to versions in other studies and recommended their use for future research. Reported internal consistency for the self-report was low for Peer Problems, Hyperactivity-Inattention, Pro-social Behavior and Conduct Problems (α = .21, .23, .42 and .48, respectively) and good for Emotional Problems (.71) and Total Difficulties (.74). Test-retest reliability was reported as acceptable for Total Difficulties (.71) and all subscales (.62 to .71) except for Peer Problems (.45). Significant correlations between parent and self-report were found for all subscales and total score (.20 to .26) except for Hyperactivity-Inattention (.18). The authors also report lower fit indices on the CFA of

the 5-factor model for the self-report (CFI: .718; TLI: .680 and RMSEA: .071) than the parent- and teacher-report versions.

Challenges in assessing DHH children

Van Gent, Goedhart, and Treffers (2012) reported that DHH adolescents were significantly older at their first referral than their TH peers and emphasize the need for preventive interventions for early recognition of mental health problems. To ensure early recognition and valid assessment Ohre, Saltnes, von Tetzchner, and Falkum (2014) and Roberts et al. (2015) emphasize the need for instruments in sign languages. There is, however, a considerable lack of translated and validated versions of instruments such as the SDQ or the Achenbach System of Empirically Based Assessment (ASEBA; Achenbach and Rescorla, 2001) commonly used for assessing TH children. Bridging such gaps is necessary for understanding the inconsistent findings regarding the prevalence of mental health problems among DHH children. The Youth Self-Report and the SDQ have been translated to Australian Sign Language (Auslan). On both measures, DHH signing adolescents have reported more difficulties on the Auslan than the written versions (Cornes & Brown, 2012; Cornes, Rohan, Napier, & Rey, 2006).

Except for the pilot study by Aanondsen, Heiling, Nøvik, and Jozefiak (2018), there are hardly any studies on Norwegian DHH children's mental health and no studies on the validation of assessment tools in NSL for assessing mental health in DHH children. Norway is unique in offering parents of DHH children 40 weeks (i.e. 2-4 weeks/year) of NSL classes over 16 years with all expenses covered. Therefore, one might expect a higher level of signing skills amongst Norwegian DHH children and their parents. This, in turn, may have a positive influence on their mental health. As some studies have found that DHH adolescents report more symptoms on assessments based on sign language (Cornes & Brown, 2012;

Cornes et al., 2006), validation studies on assessment tools in NSL are necessary. The present study provides psychometric properties for the Norwegian version of the SDQ self-report (SDQ-NOR), which is the first instrument translated to Norwegian Sign Language (NSL) for assessing mental health in children.

Aims

The main aims of this study were to validate the SDQ self-report in NSL (SDQ-NSL) and to establish the psychometric properties of the SDQ-NOR, as previous studies have shown marked differences in the prevalence of mental health problems based on written versus signed instruments (Brown & Cornes, 2015; Cornes & Brown, 2012; Cornes et al., 2006). The usability of the SDQ-NSL for signing DHH children was assessed from the children's perspective. Finally, rates of emotional and behavioral problems as classified by Norwegian cut-off scores were examined for Norwegian DHH children based on both the written and the NSL versions of the SDQ self-report.

We addressed the following research questions:

1. What are the psychometric properties (internal consistency, test-retest reliability and construct validity) of the SDQ-NSL and SDQ-NOR for DHH children?
2. What are the correlations between the total score, subscales, and items between the SDQ-NSL and SDQ-NOR?
3. What are the correlations between the total score, subscales, and items between the self-report (SDQ-NSL and SDQ-NOR) and parent-report?
4. What do DHH children think about the usability of the SDQ-NSL and SDQ-NOR?

5. What are the rates of emotional and behavioral problems in the clinical range based on the SDQ-NSL and the SDQ-NOR?

Methods

Participants

Caluraud et al. (2015) reported that hearing loss (HL) of >40 dB affects 1.4 of every 1000 infants (mild HL in 13%, moderate HL in 50%, severe HL in 17% and profound HL in 20%). In Central and Northern Norway, this amounts to an estimate of 205 children with a HL of >40 dB, i.e. 35 with severe and 41 with profound HL, based on a population of 146.308 children aged 9 to 18 years. For the whole country, this amount to an estimate of 151 with a severe and 177 with a profound HL, based on a population of 633.295 children aged 9 to 18 years.

DHH children aged 9 to 17 years who were enrolled part- or full time at A.C. Møller School, a school for deaf children of Central and Northern Norway, for the school year of 2016/17 were invited to participate. DHH adolescents aged 15 to 20 years attending upper secondary school in Central Norway with NSL as their first or second language were also invited. The overall response rate was 86% (49/57). Parents (from the mainstream and the deaf school) also took part in the study.

<figure 1 here>

Two children were excluded based on the assessment of their deaf school teacher because they lacked fluency in NSL as they only recently had started learning NSL. Apart from fluency in both written Norwegian and NSL, we applied no exclusion criteria. Forty-nine DHH children, 35 of them girls (71.4%), participated in this study. The mean age was 13.5 years (SD = 2.99; range = 9-20) and the mean nonverbal IQ was 108.9 (SD = 18.1;

range = 49-143) based on cognitive assessment with Leiter 3. Twenty-nine of 32 (90.6%) mothers had completed 12 years or more of education, whereas 23 of 31 (74.2%) fathers had completed 12 years or more of education. Data were collected between November 1, 2016, and May 9, 2017. The majority of the DHH children (65.6%) mainly attended mainstream schools and spent 2 to 6 weeks at the deaf school per school year.

Hearing- and language-related information for the participants in this study can be found in Tables 1 and 2.

<Tables 1 and 2>

Measures

Sociodemographic and hearing related information. Parents completed a questionnaire about their children's age, sex, socioeconomic status, and parents' physical and mental health. The parents also completed a questionnaire developed for and used in a pilot study for assessing type and severity of hearing loss, type of schooling, and parents' attendance at sign language classes.

Language-related information.

Parents were asked to respond to questions about their children's preferred mode of communication (spoken Norwegian, NSL, other spoken or sign language or bilingual) within and outside the family.

Spoken Language Skills. The participants' auditory performance (speech intelligibility and listening skills) was assessed by parents using the Categories of Auditory Performance (CAP; Archbold, Lutman, & Marshall, 1995) and the Speech Intelligibility Rating (SIR; Allen, Nikolopoulos, Dyar, & O'Donoghue, 2001). CAP and SIR are frequently used in research. The CAP is a single-item scale with a range of 0 to 7. Level 0 is ''no

awareness of environmental sounds'' and Level 7 ''uses a telephone with a known speaker.'' The SIR is also a single-item scale and has a range of 1 to 5. Level 1 is ''connected speech is unintelligible.'' and 5 ''connected speech is intelligible to all listeners''. Interrater reliability for the Danish version was reported as good (CAP: kappa = .785; SIR: kappa = .848; Dammeyer, 2010b). The sum of CAP and SIR was calculated for each child as the spoken language skills score.

Sign Language Skills. The participants' sign language skills were assessed with the Norwegian version of the Sign Language Production Scale (SPS) and the Sign Language Understanding Scale (SUS) developed by Dammeyer (2010b). SPS and SUS were designed by Dammeyer (2010b) as a short screening of sign language skills for research purposes. The structure and range of the SUS and SPS corresponds to that of the CAP and SIR scales. The SPS is a single-item scale with a range of 1 to 5. Level 1 is ''the child does not produce real signs'' and Level 5 ''the child uses fluent and almost conventional correct sign language.'' The SUS is a single-item scale with a range of 0 to 7. Level 0 is ''does not react to or does not comprehend signs'' and Level 7 ''is able to take in long and complex conversations in sign language.'' The interrater reliability of the Danish version was reported to be good (kappa = .944 for SUS and .921 for SPS; Dammeyer, 2010b). The validity of the Danish version of the SUS was evaluated by comparing the ratings of 12 children with their scores on the Danish translation (Seiler & Larsen, 2005) of the Assessing British Sign Language Development: Receptive Skills Test (Herman, Holmes, & Woll, 1999). The correlation between the SUS and the Receptive Skill Test reached statistical significance (Spearman rank correlation coefficient = .905, $p < .001$; Dammeyer, 2010a). No corresponding test was available for sign language production. The ''sign language skills score'', was calculated for each child by summing the SPS and SUS scores.

Leiter International Performance Scale – Third Edition (Leiter-3)

The nonverbal intelligence of the participants was assessed using the following subtests of Figure Ground, Form Completion, Classification/Analogies and Sequential Order from the Leiter-3. The composite score for nonverbal intelligence is based on the sum of the scaled scores for these subtests (Roid, Miller, Pomplun, & Koch, 2013).

Emotional and behavioral problems

The SDQ (R. Goodman, 1997) is a multi-informant mental health assessment. For this study, we administered both the parent and the self-report of the SDQ. Each version of the questionnaire comprises twenty-five questions, each scored on a three-point Likert scale (0 = "Not true," 1 = "Somewhat true" and 2 = "Certainly true"). These questions can be divided into five subscales measuring Emotional Problems, Conduct Problems; Hyperactivity-Inattention, Peer Problems and Pro-social Behavior, as well as a Total Difficulties scale of overall psychological adjustment based on the four negative subscales, with higher scores indicating more difficulties.

The SDQ self-report was originally designed for adolescents aged 11 to 16 years (R. Goodman, 2001). Muris, Meesters, Eijkelenboom, and Vincken (2004), however, suggested that the self-report may also be used for children as young as 8, whereas van Roy et al. (2006) find evidence that it can be used for adolescents as old as 19. Based on this evidence of acceptable psychometric properties for both younger and older children as well as the need for assessment tools of mental health in NSL for children of all ages we have included children aged 9 to 20. In our study, children completed both the written and signed self-report versions of the SDQ.

The translation process

We based the translation process of the SDQ on the guidelines for cross-cultural adaptation of written self-report measures by Beaton, Bombardier, Guillemin, and Ferraz (2000) as well as on adaptations suggested by Roberts et al. (2015) based on differences in syntax, morphology and prosody of sign languages and their visual nature. Two independent forward and backward translations of all scales of the SDQ from written Norwegian to NSL were completed. The forward translations were conducted and filmed by two bilingual deaf native NSL users with university degrees in teaching. A panel consisting of the translators, a clinical psychologist, a colleague with a graduate degree in medicine specializing in child and adolescent psychiatry and a consultant with a master's degree in language and communication and fluent in NSL discussed semantic, conceptual, lexical and cultural differences and developed a consensus-based forward translation that was filmed. This forward translation was then presented to a focus group consisting of teachers from the local deaf school. The teachers (deaf, hearing and CODA, i.e. a typically hearing person raised by deaf parents), were asked to evaluate whether DHH children with a mixture of language experiences and levels of fluency would be able to understand the translation. The consensus version was adjusted according to the feedback of the focus group and filmed again. Two independent backward translations of the final consensus version were conducted by two hearing sign language interpreters, one of them with a background as CODA and a master's degree in language and communication. The backward translations were reviewed by the panel and compared to the original written Norwegian version.

To gain approval from YouthInMind (SDQ's copyright holders) the Norwegian back translation was then translated to English. YouthInMind approved items and made suggestions for those not approved. These went back through the translation cycle until final approval was achieved. After the final approval, the SDQ-NSL was filmed professionally and prepared for interactive online administration using Select Survey.

Procedures

Enrolled children and their parents received verbal and written information about participating in the study during their first attendance at the school after the study was initiated. Written informed consent was obtained from adolescents and parents prior to inclusion. The participating children responded to the web-based SDQ-NSL and SDQ-NOR as well as a question about the usability of the two versions and completed a nonverbal cognitive assessment. The administration of the SDQ-NSL and SDQ-NOR was conducted on two separate occasions, two days apart. The order of these two administrations was randomized. The same procedure was applied for collecting retest data when the children returned for their next stay at the deaf school about 15 weeks after the first data collection. DHH children had access to their teacher and a psychologist, who were both fluent in sign language, during data collection. When children asked for help with the SDQ-NSL they received help in NSL, whereas children responding to the SDQ-NOR were assisted in spoken Norwegian or sign supported speech.

Statistical analyses

Seventeen of the 49 parent-reports (34%), nine of the 49 SDQ-NSL self-reports (18.4%) and three of the 49 SDQ-NOR reports were not completed. These missing cases were excluded from the analyses. There were no missing items as YouthInMind requires a response to all items on the SDQ in web-based administrations. Five of 14 adolescents aged 16 years or older (35.7%) consented to their parents' participation in the study; two of the five parents (40%) completed the parent-report (see Fig. 1). On average, adolescents who consented to their parents' participation reported a lower total score on the self-reported SDQ than those who did not consent, but the differences were not statistically significant.

To examine boys' and girls' mean score differences, two-sample t-tests were calculated for the four subscales and for the total score.

Dillon−Goldstein's rho (DG rho) was used to assess internal consistency because of the limitations of Cronbach's α, such as assumptions of uncorrelated errors, tau-equivalence, and normality (Yanyun & Green, 2011). DG rho was interpreted as acceptable at .6 to .7, and as good when > .7.

Test-retest reliability based on intraclass correlations (ICC) was calculated using a two-way random effects model. ICC values of less than .5 were considered poor, .5−0.75 was acceptable, .75−0.9 was good, and greater than .90 was considered excellent reliability (Koo & Li, 2016).

Spearman's rank correlation coefficient was used to compare total score, subscales and items between the SDQ-NSL and SDQ-NOR. Significant differences between scores on subscales, and total scores were established based on paired t-tests.

We conducted a confirmatory factor analysis (CFA) with the weighted least squares means and variances (WLSMV) estimation method for categorical variables to confirm the original factor structure of the SDQ (five-factor model) for DHH children for the SDQ-NOR as well as for the SDQ-NSL. Further CFAs were carried out for the one-factor model, the three-factor model as well as the 2nd order model. The chi-squared test, the normed chi-square (χ^2/df), the root mean square error of approximation (RMSEA), comparative fit index (CFI) and Tucker-Lewis Index (TLI) were used to assess model fit. A non-significant chi-square test, CFI and TFI > .9, RMSEA < .1 were considered indicators of acceptable goodness of fit according to Mehmetoglu and Jakobsen (2017), whereas CFI and TFI > .95 and RMSEA <.05 were considered as indicators of good model fit (Hu & Bentler, 1999). A normed chi-square of <2.0 was considered acceptable for this study although others have reported

acceptable ratios as high as 5.0 (Hooper, Coughlan, & Mullen, 2008). Standardized factor loadings greater than .4 were considered acceptable (Mehmetoglu & Jakobsen, 2017).

As Hair, Hult, Ringle, and Sarstedt (2016) point out, small sample size can cause problems with under-identified models and non-convergence in CFA. The estimator WLSMV has been shown to overestimate interfactor correlations when the sample size is relatively small (Li, 2016). Partial least squares structural equation modeling (PLS-SEM) has been shown to be less prone to these problems as it is nonparametric and makes fewer distributional assumptions. PLS-SEM, however, is mostly used for exploratory purposes as it lacks goodness of fit measures. Because of the small sample size, we also carried out PLS-SEM to establish factor loadings and discriminant validity (average variance extracted (AVE)) as suggested by Hair et al. (2016). Factors with AVE scores greater than .5 were regarded as satisfactory for convergent/discriminant validity. Fornell and Larcker (1981), however, argue that AVE >.4 can be treated as acceptable as long as composite reliability is above .6.

We conducted Spearman's rank correlations to assess multi-informant correlations between the parent- and both the self-reported (NSL and NOR) scaled scores of the SDQ. These were compared with multi-informant correlations described in other samples by Achenbach et al. (1987), R. Goodman (2011) and Roberts et al. (2015).

A contingency table was computed for comparing the total score of DHH children within the normal, sub-clinical and clinical ranges for the SDQ-NSL and SDQ-NOR, as well as the concordance between the two self-reports. Based on Fagerland, Lydersen, and Laake's (2017) recommendation we computed a contingency table and used Fisher's exact test to examine the association between the DHH children's preferred mode of communication in everyday life and their preference for the SDQ-NSL or SDQ-NOR. Further contingency tables including Kruskal-Wallis equality-of-populations rank tests were computed for the

DHH children's spoken as well as their NSL skills and their preference for the SDQ-NSL or SDQ-NOR.

Descriptive analyses were conducted in IBM SPSS Statistics version 25, the CFA was carried out in MPlus version 8 and ICC, two-sample and paired sample t-tests, DG rho and Spearman rank correlations were conducted in Stata/SE 14.2 for Windows. PLS-SEM including AVE was conducted in Stata applying the module for PLS-SEM by Venturini and Mehmetoglu (2017). For all analyses, alpha levels of <.05 were considered statistically significant.

Ethics

Written informed consent was obtained from all parents and from adolescents older than 16 years prior to inclusion. Verbal informed consent was obtained from children under the age of 16 years. Study approval was given by the Regional Committees for Medical and Health Research Ethics (reference number: 2015/1739/REK midt).

Results

Table 3 presents the means and standard deviations for the DHH participants on the self-report of the SDQ (SDQ-NSL and SDQ-NOR).

<Table 3 here>

A two-sample t-test of the girls' and boys' mean scores on the five subscales and total scores showed no significant difference in gender for either of the self-report versions.

Reliability

Internal consistency

<Table 4 here>

As can be seen in Table 4, internal consistency was found to be acceptable to good for all subscales for both the SDQ-NSL and SDQ-NOR based on DG rho.

Test-retest reliability

Test-retest data were collected for 20 participants after they returned to the deaf school. An average of 15.17 weeks (SD = 1.01) elapsed between T1 and T2 for the SDQ-NSL and an average of 15.03 weeks (SD = 1.05) elapsed for the SDQ-NOR. Test-retest correlations are presented in Table 5.

<Table 5 here>

As can be seen from Table 5, test-retest reliability for SDQ-NSL was found to be acceptable based on ICC for all subscales as well as Total Difficulties. Test-retest reliability for the SDQ-NOR was established as acceptable for Emotional Problems, Hyperactivity-Inattention and Peer Problems, as good for Emotional Problems and Total Difficulties, and as poor for Pro-social Behavior.

Validity

Construct validity

The standardized factor loadings, AVE as well as model fit indices for the subscales of the five-factor model are displayed in Table 6 (SDQ-NSL) and Table 7 (SDQ-NOR).

<Tables 6 and 7 here>

The goodness of fit indices indicated a better fit for the SDQ-NSL than the SDQ-NOR for the DHH children in this study as the SDQ-NSL showed acceptable model fit on two indices ($\chi 2$/df and RMSEA), and the SDQ-NOR showed acceptable fit on one ($\chi 2$/df). Factor loadings based on CFA and PLS-SEM were acceptable for 4-5 items of the Emotional

Problems, Conduct Problems and Hyperactivity-Inattention subscales for both the SDQ-NSL and the SDQ-NOR. For details on the negative factor loading of item 11 of the SDQ-NSL displayed in Table 8, see Appendix A. The subscales of Conduct and Peer Problems showed an interfactor correlation of 1.053 on the SDQ-NOR. None of the modification indices for SDQ-NSL or SDQ-NOR suggested correlated residuals for the five-factor model.

AVE was above the acceptable .5 only for Emotional Problems on the SDQ-NOR. Fornell and Larcker (1981), however, argue that AVE > .4 can be treated as acceptable as long as composite reliability, in this case DG's rho, is above .6. This was the case for Emotional Problems and Hyperactivity-Inattention on the SDQ-NSL and for Conduct Problems, Hyperactivity-Inattention and Pro-social Behavior on the SDQ-NOR.

<Table 8 here>

Further comparison of the structure of the SDQ-NSL and SDQ-NOR (see Table 8) showed that the data failed to satisfy the strictest interpretation of goodness of fit measures for the five-factor model as well as the one-factor model (SDQs total score based on the four problem scales). CFA of the proposed three-factor model (A. Goodman et al., 2010) did not converge for either the SDQ-NSL or the SDQ-NOR for DHH children. Overall, the SDQ-NSL showed more acceptable fit than the SDQ-NOR for DHH children in this study for both the five-factor, one-factor, and second-order models. When comparing these different factor models, the SDQ-NSL demonstrated best fit for the five-factor and one-factor model with acceptable fit on two ($\chi 2$/df and RMSEA) of the five goodness of fit measures.

Comparison of SDQ-NSL and SDQ-NOR

To compare the SDQ-NSL with the SDQ-NOR self-report, Spearman rank correlations were calculated for the five subscales and the total score (Table 9).

<Table 9 here>

All the correlations were highly significant at $p < .001$. The correlations for Emotional Problems, Peer Problems, and Total Difficulties were in the good range, while Conduct Problems, Hyperactivity-Inattention and Pro-social Behavior demonstrated acceptable correlations. DHH children reported significantly more Conduct Problems on the SDQ-NSL ($M = 2.35$, $SD = .301$) than on the SDQ-NOR ($M = 1.55$, $SD = .286$); $t(39) = 3.439$, $p = .001$, but other differences were not significant.

All items for Emotional, Conduct and Peer Problems were significantly correlated for the two versions, mostly moderately to strongly (.323 to .736). All items on the Hyperactivity and Inattention subscale were significantly correlated (weak to moderate correlations; .277 to .535), apart from item 10, "fidgety". The items on the Pro-social Behavior subscale were not significantly correlated (.102 to .371), apart from item 9, "caring".

Multi-informant correlations

Multi-informant correlations between the scores of DHH children and their parents on the self- report SDQ-NSL and SDQ-NOR are presented in Table 10. Correlations between the self- and parent-report were significant for Emotional Problems for both the SDQ-NSL and SDQ-NOR.

<Table 10 here>

Symptom levels

The number of DHH children classified as reporting symptoms in the normal, sub-clinical and clinical ranges on the SDQ was calculated based on Norwegian cut-off scores (Rønning et al., 2004) and is shown in Table 11.

<Table 11 here>

Five DHH children (12.5%) were classified as Clinical on both the SDQ-NSL and the SDQ-NOR (concordance of 71.4%). Based on both scales, seven of 40 DHH children (17.5%) were identified with symptoms in the clinical range.

Usability

When asked which version of the SDQ the DHH children preferred, 44.9% (22/40) preferred the SDQ-NOR. The SDQ-NSL or a combination of the signed and written self-report was the preferred choice of 30.6% (15/40) and 6.1% (3/40) did not know.

During administration of the SDQ-NSL and SDQ-NOR, children commented on the fact that they spent more time completing the SDQ-NSL as it took longer to view the video clips of the signed items than to read the items. Based on Fisher's exact test no significant association was found between the children's preferred mode of communication in everyday life (based on parent-report) and the children's preference for the signed or written versions of the SDQ. In addition, no significant associations between the parents' assessment of their children's spoken (CAP and SIR) and sign language skills (SUS and SPS) and the children's preference for the NSL or written version were found based on Kruskal Wallis tests (see Appendix B).

Discussion

Internal consistency and test-retest reliability were established as acceptable to good. CFA resulted in the best fit for the proposed five-factor model for both versions, although not all fit indices reached acceptable levels. The SDQ-NSL and SDQ-NOR both demonstrated similar psychometric properties to those reported for the SDQ in other studies both for TH (Essau et al., 2012; R. Goodman, 2001; van Roy et al., 2008) and DHH children (Cornes & Brown, 2012; Hintermair, 2007; Niclasen & Dammeyer, 2016; Roberts et al., 2015), except for the subscale of Pro-social Behavior on the SDQ-NSL.

Both self-report versions demonstrated acceptable levels of internal consistency although the DG coefficients were higher than Cronbach's α reported in other studies on both DHH (SDQ-Auslan: α = .42 to .83; SDQ-BSL: α = .21 to .74) and TH children (Cornes & Brown, 2012; Essau et al., 2012; R. Goodman, 2001; Roberts et al., 2015; Viana, Rabian, & Beidel, 2008). A possible explanation for this may be the known tendency of Cronbach's α to underestimate internal consistency due to its limitations (assumptions of uncorrelated errors, tau-equivalence, and normality; Yanyun & Green, 2011). Acceptable but relatively lower internal consistency was found for Peer Problems (DG = .68 compared to .74 to .88 on the other scales) on both the SDQ-NSL and SDQ-NOR. This is in accordance with other studies on DHH with a Cronbach's α of .42 (Peer Problems) compared to .55 to .83 for the Auslan version and a Cronbach's α of .21 (Peer Problems) compared to .23 to .74 for the; BSL (Cornes & Brown, 2012; Roberts et al., 2015). For the SDQ-NSL the lowest, but still acceptable, internal consistency was found for Pro-social Behavior.

The interval of 15 weeks between test and retest was too long to be regarded as a good measure of test-retest reliability. Score differences may reflect actual changes over time, so correlations here should be seen as a lower bound of test-retest reliability. As data collection was dependent on the participants stay at the deaf school it was not possible to shorten the

interval even though this would have been desirable. Cornes and Brown (2012) reported higher test-retest reliability for the SDQ-Auslan (.75 to .85), this, however, was based on an interval of two days only. The mean test-retest correlation for the SDQ-NSL (.62) was similar to that found by R. Goodman (2001) for a test-retest interval of 4-6 months (.62), as well as that reported by Roberts et al. (2015) for the SDQ-BSL for an interval of three weeks (.61). Overall, test-retest reliability was established as acceptable for both self-report versions.

Correlations between the two self-report versions were all significant except for Pro-social Behavior, and much higher than those reported by Cornes and Brown (2012). This may indicate closer correspondence between the SDQ-NSL and SDQ-NOR, either due to more equivalent phrasing in both written Norwegian and NSL, greater literacy or the high number of children with a spoken language preference among this DHH sample. Literacy, however, was not assessed in this study, therefore it is difficult to conclude on this subject. Examination of the interitem correlations for the two versions showed that items 2 (“I am restless. I cannot stay still for long”) and 10 (“I am constantly fidgeting or squirming”) were not significantly correlated. Rønning et al. (2004) have previously described the semantic similarity between these two items in spoken/written Norwegian based on CFA. In NSL, however, the items are more distinct, which may indicate that the SDQ-NSL is better able to differentiate between the two items. The inter-item correlations for four of the five items on Pro-social Behavior were not significant. The non-significant correlations of Pro-social Behavior at item level may be an indication of an issue with the translation of these items. The items on the SDQ-NSL for Pro-social Behavior should, therefore, be evaluated by new forward and back translations and reviewed by a new reference group.

Other studies have reported problems with the internal consistency of the Conduct and Peer Problem scales for the SDQ self-report (Cornes & Brown, 2012; Essau et al., 2012; R. Goodman, 2001; van Roy et al., 2008). The same pattern can be seen for both the SDQ-NSL

and SDQ-NOR based on the results for discriminant validity (AVE). A possible explanation for this phenomenon on the self-report may be that children's answers are influenced by their knowledge of the social desirability of positive social behavior and they are therefore less likely to admit negative behavior or problems with peers than their teachers or parents. Another explanation may be that the children's understanding of conduct and peer problems are closely linked, and therefore the factors are also correlated.

In the comparison of several different factor models for the SDQ-NSL and SDQ-NOR (five-factor, one-factor, three-factor, and 2nd-order models) the SDQ-NSL demonstrated the best fit for the five-factor model with an acceptable fit on two ($\chi 2$/df and RMSEA) of the five goodness of fit measures. Studies including all three informants have previously shown that the parent and teacher versions of the SDQ show better model fit than the self-report in both TH and DHH children (R. Goodman, 1997; Roberts et al., 2015). The fit indices for the SDQ-NSL (CFI: .747, TLI: .715 and RMSEA: .107) are similar to those reported in the BSL study (CFI: .718; TLI: .680 and RMSEA: .071). It should also be noted that the three-factor model did not converge for either version, which is consistent with Norwegian validation studies (Rønning et al., 2004; van Roy et al., 2006; van Roy et al., 2008) and Essau et al.`s (2012) findings for Sweden, the UK, and Germany. It should be noted, however, that the small sample size might have contributed to non-convergence of the three-factor model. As none of the participants answered "Not true" on item 11 ("I have one good friend or more") of the SDQ-NSL, the empty cells caused the negative factor loading for that item. A larger sample is likely to have secured an answer for all alternative categories. The interfactor correlation of greater than 1 between Conduct and Peer Problems on the SDQ-NOR can be explained by Li's (2016) findings that the WLSMV estimator demonstrates a tendency to overestimate interfactor correlations in small sample sizes. The non-convergence of the three-factor model for both self-report versions, the negative factor loading on item 11 of the SDQ-NSL and the

interfactor correlation greater than 1 for Conduct and Peer Problems on the SDQ-NOR do leave some uncertainty regarding the correct identification of the CFA models. As the goodness of fit indices of the CFAs in this study were similar to those in other studies (R. Goodman, 2001; Roberts et al., 2015) it is, however, likely that a larger sample would confirm our present results.

Multi-informant correlations for both self-report versions were close to the mean of .25 reported by Achenbach et al. (1987) in their meta-analysis. Multi-informant correlations (total score and subscales) for the SDQ-NSL (range .03 to .32) were similar to those found for the SDQ in British Sign Language, (range .18 to .26) (Roberts et al., 2015) although lower than those reported by R. Goodman (2001). The parent-child correlations for Emotional Problems, however, were significant for both versions and greater than the mean correlation of .25 reported by both Achenbach et al. (1987) and Cornes & Brown (2012) for the Auslan version. A possible explanation for this may be the easy access to early intervention as well as sign language tuition for parents of DHH children in Norway, which may, in turn, lead to better communication skills about emotions between parents and DHH children. Laugen, Jacobsen, Rieffe, and Wichstrom (2017) found that parents of preschool children with hearing loss were more accurate in estimating their child's emotion understanding than parents of TH children. The parents' more accurate estimation of their children's emotion understanding (Laugen et al., 2017) may have contributed to the higher level of agreement on emotional problems found in this study

The majority of the DHH children reported preferring the SDQ-NOR (44.9%). Analyses showed that language preference in everyday life as well as level of spoken and sign language skills did not influence the participants' preference for the written SDQ or SDQ-NSL. A possible explanation for this is that the children's everyday language preferences are parent-reported as the children themselves were not asked to report their

language preferences for everyday life. The children's spontaneous feedback during administration indicated that the preference for the written version was related to the more time-consuming nature of the video presentation of the NSL version. It is, however in contrast to studies reporting reading difficulties in many DHH children (Harris et al., 2017; Marschark, Rhoten, & Fabich, 2007; Marschark et al., 2009). It is possible that the high correlations between the SDQ-NSL and SDQ-NOR in our sample are due to better literacy in this sample than in the one reported by Cornes and Brown (2012). This, in turn, could explain the preference for the SDQ-NOR. As we only assessed spoken and sign language skills but not literacy, this cannot be tested within our study.

The higher level of emotional and behavioral symptoms reported by DHH children on the SDQ-NSL for most subscales is in accordance with Cornes and Brown's (2012) and Brown et al.'s (2006) findings. The rate of emotional and behavioral symptoms in the clinical range for DHH children (17.5% on both self-report versions) was almost twice as high as that reported in a Norwegian community sample (Rønning et al., 2004). This is in accordance with other studies reporting an elevated prevalence of emotional and behavioral problems in DHH children, but somewhat lower than the 20−50% found in other studies (Fellinger et al., 2012; Stevenson et al., 2015; Theunissen et al., 2014). This is in accordance with Dammeyer's (2010b) study, which found higher rates of emotional and behavioral problems in Danish DHH children as compared to typically hearing children. The same study, however, found no increased risk of mental health problems in DHH children with good communication skills. As previously suggested by Mejstad et al. (2009) the extent of parental support provided in Nordic countries may ensure better mental health in DHH children. Mejstad et al. (2009) also found that boys in their DHH sample reported mental health similar to that of community samples, whereas girls in their sample reported significantly more emotional and behavioral problems. In our study, 71.4% of the participants were girls. No

significant differences between the reported mean scores on subscales and total score on either version were found, there was, however, a slight tendency for girls to report more symptoms on Emotional and Peer Problems, and to have a higher Total Score. The small number of boys in this sample may have contributed to a failure to replicate the findings of Mejstad et al. (2009). Girls have been found to report more symptoms on Emotional Problems in other studies as well (Rønning et al., 2004). A further possible explanation for the higher level of emotional and behavioral problems could be the broad range of cognitive abilities in this sample. A closer examination of the data, however, proved the participant with the lowest non-verbal IQ to be an extreme outlier in the IQ distribution. It is therefore not very likely to have influenced the rate of emotional and behavioral problems in this study. The rates of additional impairment in this study, learning and visual impairment, in particular, are equivalent to those reported in other studies (Armitage, Burke, & Buffin, 1995; Gallaudet Research Institute, 2008) and therefore not likely to have influenced the rate of symptoms in this study either.

The items for Pro-social Behavior on the SDQ-NSL should, however, be reevaluated using a thorough forward and backward translation process to ensure that the items measure the same concept in both versions. As Pro-social Behavior is not part of the Total Difficulties score, this does not pose a problem for the validity of the total score and the assessment of mental health problems.

Limitations

A major limitation of this study was the small sample size due to the limited number of signing DHH children in the population. The sample size here was lower than the minimum number of cases recommended for CFA and other multivariate analyses based on covariance. This, in turn, poses a problem for a thorough psychometric evaluation of the SDQ-NSL and SDQ-NOR for DHH children. To compensate for the effects of small sample size on CFA,

we used the PLS-SEM as well, which is known to be more robust for such situations (Hair et al., 2016). The combination of analyses used here was chosen as the best practical solution for this study but does still leave room for uncertainty regarding the conclusions.

A further limitation is the absence of a gold standard for establishing criterion validity for mental health problems in DHH children. The use of a written instrument such as the Youth Self-Report (ASEBA), as a gold standard would not have been reliable or valid because there is evidence that many DHH children have difficulties reading (Harris et al., 2017; Marschark, Rhoten, et al., 2007; Marschark et al., 2009). Further, the use of a verbal clinical interview without an interpreter or signed supported speech would not have been possible or valid because of the participants' level of hearing loss. In addition, there are no existing studies on the reliability and validity of the simultaneous translation of a semi-structured diagnostic interview such as the Schedule for Affective Disorders and Schizophrenia – Present Life Version 2009 (Kiddie–SADS-PL 2009; Kaufman et al., 1997) to NSL.

Conclusion

The evaluation of the psychometric properties of the self-report SDQ-NSL is promising. It primarily suffers from the same weaknesses as found in other studies of the self-report (written and signed). Questions may be raised regarding the quality of the items for Pro-social Behavior on the SDQ-NSL. The use of the SDQ-NSL for assessing mental health in DHH children may, therefore, be recommended. Based on the participants' feedback, the correspondence between the two self-report versions and their similar psychometric properties, we recommend administering the SDQ self-report with both written and signed items in a combined web-based version. As the validation is based on a small sample, further assessment of its psychometric properties in a larger sample is recommended. Further

research on DHH children is needed to ensure early detection and intervention, reliable and valid assessment, and treatment of emotional and behavioral problems. Because of the small number of signing DHH children in the population, cross-cultural studies should be encouraged. This would increase the possibility of conducting research on larger samples as well as allowing examination of cross-cultural similarities and differences.

Funding

This work was supported by grants from the Regional Hospital Trust in Central Norway; the Department of Child and Adolescent Psychiatry at St. Olavs University Hospital; the Research Department at St. Olavs University Hospital; and the Gidske and Peter Jacob Sørensens Research Foundation.

Acknowledgments

This study was a collaboration between the Department of Child and Adolescent Psychiatry at St. Olav's University Hospital, and the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU), Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Trondheim. We would like to thank STATPED and Tiller Upper Secondary School for allowing us to recruit participants as well as all children, adolescents and parents participating in this study. We would also like to thank Professor Stian Lydersen for his valuable assistance with some of the statistical analyses, and Maria Gascon Ramos for her useful feedback.

Conflict of interest

On behalf of all the authors, the corresponding author states that there is no conflict of interest. The first author has been involved in the translation of the SDQ-NSL but has not gained any economic benefits from that or retained any financial interests in the SDQ-NSL.

References

Aanondsen, C. M., Heiling, K., Nøvik, T. S., & Jozefiak, T. (2018). Quality of Life, family function and mental health of deaf and hard-of-hearing adolescents in mental health services in Norway – a pilot study. *International Journal on Mental Health and Deafness, 4*(1), 3–13. Retrieved from http://www.ijmhd.org

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*(2), 213–232. doi:10.1037/0033-2909.101.2.213 3562706

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles: child behavior checklist for ages 6-18, teacher's report form, youth self-report: an integrated system of multi-informant assessment.* Burlington, VT: ASEBA.

Allen, C., Nikolopoulos, T. P., Dyar, D., & O'Donoghue, G. M. (2001). Reliability of a rating scale for measuring speech intelligibility after pediatric cochlear implantation. *Otology & Neurotology, 22*(5), 631–633. doi:10.1097/00129492-200109000-00012

Archbold, S., Lutman, M. E., & Marshall, D. H. (1995). Categories of auditory performance. *Annals of Otology, Rhinology, and Laryngology. Supplement, 166*, 312–314. Retrieved from https://journals.sagepub.com/loi/aor

Armitage, I. M., Burke, J. P., & Buffin, J. T. (1995). Visual impairment in severe and profound sensorineural deafness. *Journal Archives of Disease in Childhood*, 73(1), 53-56. doi:10.1136/adc.73.1.53

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*(24), 3186–3191. doi:10.1097/00007632-200012150-00014

Brown, P. M., & Cornes, A. (2015). Mental health of deaf and hard-of-hearing adolescents: What the students say. *Journal of Deaf Studies and Deaf Education, 20*(1), 75–81. doi:10.1093/deafed/enu031

Caluraud, S., Marcolla-Bouchetemblé, A., De Barros, A., Moreau-Lenoir, F., De Sevin, E., Rerolle, S., Charrière, E., Lecler-Scarcella, V., Billet, F., Obstoy, M.-F., Amstutz-Montadert, I., Marie, J.-P., & Lerosey, Y. (2015). Newborn hearing screening: Analysis and outcomes after 100,000 births in Upper-Normandy French region. *International Journal of Pediatric Otorhinolaryngology*, *79*(6), 829–833. doi:10.1016/j.ijporl.2015.03.012

Cornes, A. J., & Brown, P. M. (2012). Mental health of Australian deaf adolescents: An investigation using an Auslan version of the Strengths and Difficulties Questionnaire. *Deafness & Education International, .14*(3), 161–175. doi:10.1179/1557069X12Y.0000000009

Cornes, A. J., Rohan, M. J., Napier, J., & Rey, J. M. (2006). Reading the signs: Impact of signed versus written questionnaires on the prevalence of psychopathology among deaf adolescents. *Australian and New Zealand Journal of Psychiatry, 40*(8), 665–673. doi:10.1111/j.1440-1614.2006.01866.x 16866762

Dammeyer, J. (2010a). Congenital rubella syndrome and delayed manifestations. *International Journal of Pediatric Otorhinolaryngology*, *74*(9), 1067–1070. doi:10.1016/j.ijporl.2010.06.007

Dammeyer, J. (2010b). Psychosocial development in a Danish population of children with cochlear implants and deaf and hard-of-hearing children. *Journal of Deaf Studies and Deaf Education, 15*(1), 50–58. doi:10.1093/deafed/enp024

Essau, C. A., Olaya, B., Anastassiou-Hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D., O'Callaghan, J., & Ollendick, T. H. (2012). Psychometric properties of the Strengths and Difficulties Questionnaire from five European countries. *International Journal of Methods in Psychiatric Research, 21*(3), 232–245. doi:10.1002/mpr.1364

Fagerland, M., Lydersen, S., & Laake, P. (2017). *Statistical Analysis of Contingency Tables* (1 ed.). Boca Raton, FL: Chapman & Hall/CRC.

Fellinger, J., Holzinger, D., & Pollard, R. (2012). Mental health of deaf people. *The Lancet, 379*(9820), 1037–1044. doi:10.1016/S0140-6736(11)2961143-4

Fellinger, J., Holzinger, D., Sattel, H., & Laucht, M. (2008). Mental health and quality of life in deaf pupils. *European Child and Adolescent Psychiatry, 17*(7), 414–423. doi:10.1007/s00787-008-0683-y 18810312

Fellinger, J., Holzinger, D., Sattel, H., Laucht, M., & Goldberg, D. (2009). Correlates of mental health disorders among children with hearing impairments. *Developmental Medicine and Child Neurology, 51*(8), 635–341. doi:10.1111/j.1469-8749.2008.03218.x

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research (JMR), 18*(3), 382–388. doi:10.2307/3150980

Gallaudet Research Institute (2008). *Regional and national summary report of data from the 2007-08 annual survey of deaf and hard of hearing children and youth*. Washington, DC: GRI, Gallaudet University.

Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): Data from British parents, teachers, and children. *Journal of Abnormal Child Psychology, 38*(8), 1179–1191. doi:10.1007/s10802-010-9434-x

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Child Psychology & Psychiatry & Allied Disciplines, 38*(5), 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x 9255702

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1337–1345. doi:10.1097/00004583-200111000-00015

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling* (2nd ed.). Thousand Oaks, CA: Sage.

Harris, M., Terlektsi, E., & Kyle, F. E. (2017). Concurrent and longitudinal predictors of reading for deaf and hearing children in primary school. *Journal of Deaf Studies and Deaf Education, 22*(2), 233–242. doi:10.1093/deafed/enw101

Herman, R., Holmes, S., & Woll, B. (1999). *Assessing British Sign Language development: Receptive skills test*. Coleford, England: Forest Book.

Hintermair, M. (2006). Parental resources, parental stress, and socioemotional development of deaf and hard of hearing children. *Journal of Deaf Studies and Deaf Education, 11*(4), 320–330. doi:10.1093/deafed/enl005 16809431

Hintermair, M. (2007). Prevalence of socioemotional problems in deaf and hard of hearing children in Germany. *American Annals of the Deaf, 152*(3), 320–330. doi:10.1353/aad.2007.0028 18018674

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). *Evaluating model fit: A synthesis of the structural equation modelling literature.* Paper presented at the 7th European Conference on Research Methodology for Business and Management Studies, Regent's College, London, United Kingdom.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. doi:10.1080/10705519909540118

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*(7), 980–988. doi:10.1097/00004583-199707000-00021

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012

Laugen, N. J., Jacobsen, K. H., Rieffe, C., & Wichstrom, L. (2017). Emotion understanding in preschool children with mild-to-severe hearing loss. *Journal of Deaf Studies and Deaf Education, 22*(2), 155–163. doi:10.1093/deafed/enw069 27881481

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. doi:10.3758/s13428-015-0619-7

Marschark, M., Convertino, C. M., Macias, G., Monikowski, C. M., Sapere, P., & Seewagen, R. (2007). Understanding communication among deaf students who sign and speak: A trivial pursuit? *American Annals of the Deaf, 152*(4), 415–424. doi:10.1353/aad.2008.0003

Marschark, M., Rhoten, C., & Fabich, M. (2007). Effects of cochlear implants on children's reading and academic achievement. *Journal of Deaf Studies and Deaf Education, 12*(3), 269–282. doi:10.1093/deafed/enm013

Marschark, M., Sapere, P., Convertino, C. M., Mayer, C., Wauters, L., & Sarchet, T. (2009). Are deaf students' reading challenges really about reading? *American Annals of the Deaf, 154*(4), 357–370. doi:10.1353/aad.0.0111

Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using stata*. London, England: Sage Publications Ltd.

Mejstad, L., Heiling, K., & Svedin, C. G. (2009). Mental health and self-image among deaf and hard of hearing children. *American Annals of the Deaf, 153*(5), 504–515. doi:10.1353/aad.0.0069

Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the Strengths and Difficulties Questionnaire: Its psychometric properties in 8- to 13-year-old non-clinical children. *British Journal of Clinical Psychology, 43*(Part 4), 437–448. doi:10.1348/0144665042388982

Niclasen, J., & Dammeyer, J. (2016). Psychometric properties of the Strengths and Difficulties Questionnaire and mental health problems among children with hearing loss. *Journal of Deaf Studies and Deaf Education, 21*(2), 129–140. doi:10.1093/deafed/env067

Ohre, B., Saltnes, H., Von Tetzchner, S., & Falkum, E. (2014). Psychometric properties of a sign language version of the Mini International Neuropsychiatric Interview (MINI). *BMC Psychiatry, 14*(148), 1–10. doi:10.1186/1471-244x-14-148

Roberts, S., Wright, B., Moore, K., Smith, S., Allgar, V., Tennant, A., Doherty, C., Hughes, E., Moore, D. C., Ogden, R., Phillips, H., Beese, L., & Rogers, K. (2015). Translation into British Sign Language and validation of the Strengths and Difficulties Questionnaire. *Health Services and Delivery Research, 3*(2). doi:10.3310/hsdr03020

Roid, G. H., Miller, L. J., Pomplun, M., & Koch, C. (2013). *Leiter International Performance Scale Manual (3rd ed.)*. Wood Dale, IL: Stoelting.

Rønning, J. A., Handegaard, B. H., Sourander, A., & Mørch, W. T. (2004). The Strengths and Difficulties Self-Report Questionnaire as a screening instrument in Norwegian community samples. *European Child and Adolescent Psychiatry, 13*(2), 73–82. doi:10.1007/s00787-004-0356-4

Sanne, B., Torsheim, T., Heiervang, E., & Stormark, K. M. (2009). The Strengths and Difficulties Questionnaire in the Bergen Child Study: A conceptually and methodically motivated structural analysis. *Psychological Assessment, 21*(3), 352–364. doi:10.1037/a0016317

Seiler, A. F., & Larsen, T. B. (2005). *Vurdering af børns tilegnelse af dansk tegnsprog. Forståelsestest. [Assessing Danish Sign Language development]*. Aalborg, Denmark: Døveskolernes Materialecenter.

Sinkkonen, J. (1994). Evaluation of mental health problems among Finnish hearing impaired children. *Psychiatria Fennica, 25*, 52–65. Retrieved from http://www.psykiatriantutkimussaatio.fi/index.php?id=11

Stevenson, J., Kreppner, J., Pimperton, H., Worsfold, S., & Kennedy, C. (2015). Emotional and behavioural difficulties in children and adolescents with hearing impairment: A systematic review and meta-analysis. *European Child and Adolescent Psychiatry, 24*(5), 477–496. doi:10.1007/s00787-015-0697-1 25758233

Stevenson, J., Pimperton, H., Kreppner, J., Worsfold, S., Terlektsi, E., Mahon, M., & Kennedy, C. (2017). Language and reading comprehension in middle childhood predicts emotional and behaviour difficulties in adolescence for those with permanent childhood hearing loss. *Journal of Child Psychology and Psychiatry, 59*(2), 180–190. doi:10.1111/jcpp.12803

Theunissen, S. C., Rieffe, C., Netten, A. P., Briaire, J. J., Soede, W., Schoones, J. W., & Frijns, J. H. (2014). Psychopathology and its risk and protective factors in hearing-impaired children and adolescents: A systematic review. *Journal of American Medical Association Pediatrics, 168*(2), 170–177. doi:10.1001/jamapediatrics.2013.3974

Van Eldik, T. (2005). Mental health problems of Dutch youth with hearing loss as shown on the Youth Self Report. *American Annals of the Deaf, 150*(1), 11–16. doi:10.1353/aad.2005.0024

Van Eldik, T., Treffers, P. D. A., Veerman, J. W., & Verhulst, F. C. (2004). Mental health problems of deaf Dutch children as indicated by parents' responses to the Child Behavior Checklist. *American Annals of the Deaf, 148*(5), 390–394. doi:10.1353/aad.2004.0002

Van Gent, T., Goedhart, A. W., Hindley, P. A., & Treffers, P. D. A. (2007). Prevalence and correlates of psychopathology in a sample of deaf adolescents. *Journal of Child Psychology and Psychiatry, 48*(9), 950–958. doi:10.1111/j.1469-7610.2007.01775.x 17714380

Van Gent, T., Goedhart, A. W., & Treffers, P. D. A. (2012). Characteristics of children and adolescents in the Dutch national in- and outpatient mental health service for deaf and hard of hearing youth over a period of 15 years. *Research in Developmental Disabilities, 33*(5), 1333–1342. doi:10.1016/j.ridd.2012.02.012

Van Roy, B., Groholt, B., Heyerdahl, S., & Clench-Aas, J. (2006). Self-reported strengths and difficulties in a large Norwegian population 10-19 years: Age and gender specific results of the extended SDQ-questionnaire. *European Child and Adolescent Psychiatry, 15*(4), 189–198. doi:10.1007/s00787-005-0521-4

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence

Journal of Child Psychology and Psychiatry, 49, 1304–1312. doi:10.1111/j.1469-7610.2008.01942.x

Venturini, S., & Mehmetoglu, M. (2017). PLSSEM: Stata module to implement partial least squares structural equation modelling (PLS-SEM). *https://EconPapers.repec.org/RePEc:boc:bocode:s458291*.

Viana, A. G., Rabian, B., & Beidel, D. C. (2008). Self-report measures in the study of comorbidity in children and adolescents with social phobia: Research and clinical utility. *Journal of Anxiety Disorders, 22*(5), 781–792. doi:10.1016/j.janxdis.2007.08.005

Yanyun, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment, 29*(4), 377–392. doi:10.1177/0734282911406668

Figure 1 Flow-chart for the inclusion of participants (children and parents)

Table 1 Hearing-Related Characteristics (Parent-Report)

	N = 32	**%**
DHH family member(s)		
Yes/No	16/16	50.0/50.0
Time in deaf school		
1−2 days a week*	5	15.6
5 days a week	4	12.5
2−6 weeks a year*	21	65.6
> 7 weeks a year*	6	18.8
Etiology of hearing loss		
Acquired	4	12.5
Hereditary/at birth	26	81.3
Unknown	1	3.1
Missing	1	3.1
Severity of hearing loss		
Moderate: 40−70 dB	7	21.9
Severe: 71−100 dB	10	31.3
Profound: 101+	10	31.3
Unknown	5	15.6
Use of hearing aid		
CI	16	50.0
Hearing aid	24	75.0
Missing	1	3.1
Age at diagnosis		
0−2 years	19	59.4
3−5 years	13	40.6
Preferred Language		
Oral	16	50.0
Sign	7	21.9
Bilingual	9	28.1
Additional impairment		

Vision	12	37.5
Learning	3	9.4
Motor	1	3.1
Other	2	6.3
Missing	2	6.3

*Children attend both mainstream and deaf school

Table 2 Language-Related Information Based on Parent-Report

	N	**M (SD)**
Sign language skills (1-12)	28	9.46 (2.05)
Missing	4	
Spoken language skills (1-12)	30	11.37 (1.35)
Missing	2	

Table 3 Descriptive Summary of the Self-Report SDQ Scores (SDQ-NSL and SDQ-NOR): Mean and SD

SDQ scale	Emotion	Conduct	Hyperactivity	Peer Problems	Pro-social	Total score
SDQ-NSL (N = 40)	4.40 (2.41)	2.35 (1.90)	3.83 (2.40)	2.78 (1.73)	8.18 (1.46)	13.35 (6.28)
SDQ-NOR (N = 46)	4.02 (2.62)	1.61 (1.77)	3.65 (2.28)	2.89 (1.72)	8.02 (1.99)	12.17 (6.59)

SDQ Strengths and Difficulties Questionnaire possible range of score 0-40 for total score and 0−10 for each subscale
Norwegian cut-off scores (≥ 90 percentile): Emotion = 6, Conduct = 5, Hyperactivity = 7, Peer Problems = 5, Pro-social = 4, Total score = 18
SDQ-NSL: SDQ self-report in Norwegian Sign Language
SDQ-NOR: SDQ self-report in written Norwegian

Table 4 Internal Consistency Based on Dillon−Goldstein's rho for the five Subscales of the SDQ-NSL and SDQ-NOR

SDQ scale	Emotion	Conduct	Hyperactivity	Peer Problems	Pro-social
SDQ-NSL (N = 40)	.800**	.736**	.820**	.680*	.641*
SDQ-NOR (N = 46)	.876**	.780**	.798**	.682*	.825**

* acceptable internal consistency
** good internal consistency

Table 5 Intraclass Correlations and p-values for Test-retest Reliability for SDQ-NSL and SDQ-NOR

SDQ-NSL (N = 18)	ICC*	SDQ-NOR (N = 19)	ICC*
Emotional Problems	.644	Emotional Problems	.796
Conduct Problems	.649	Conduct Problems	.876
Hyperactivity	.559	Hyperactivity	.748
Peer Problems	.660	Peer Problems	.687
Pro-social Behavior	.505	Pro-social Behavior	.433
Total Difficulties	.709	Total Difficulties	.896

*all intraclass correlations were found to be significant (p-values between .011 and .001)

Table 6 Factor loadings, AVE and Model Fit Indices of the SDQ-NSL based on CFA and PLS-SEM of the Five-Factor Model.

Subscale and items	λ (CFA)	λ (PLS)	*AVE*	$\chi^2(df)$	*p*	χ^2/df	CFI	TLI	RMSEA	90% CI RMSEA
				323.766(267)	.010	1.213	.801	.776	.073	.038-.100
Emotional			.448							
3. Somatic	.474	.560								
8. Worries	.283	.618								
13.Unhappy	.760	.785								
16. Clingy	.465	.698								
24. Afraid	.911	.664								
Conduct			.383							
5. Tantrum	.831	.707								
7. Obedient	.111	.245								
12. Fights	.635	.797								
18. Lies	.653	.698								
22. Steals	.498	.484								
Hyperactivity			.482							
2. Restless	.728	.755								
10. Fidgety	.836	.734								
15. Distracted	.911	.815								
21. Reflects	.437	.588								

25. Attends	.386	.541	
Peer			.352
6. Loner	.775	.717	
11.Friend	-.358	.196	
14. Popular	.155	.210	
19. Bullied	.775	.893	
23. Oldbest	.544	.606	
Pro-social			.318
1. Considerate	.512	.253	
4. Shares	.082	.105	
9. Caring	.887	.819	
17. Kind	.398	.701	
20. Helpout	.512	.593	

Table 7 Factor Loadings, AVE and Model Fit Indices for the SDQ-NOR Based on CFA and PLS-SEM of the Five-Factor Model.

Subscale and items	λ (CFA)	λ (PLS)	*AVE*	$\chi^2(df)$	*p*	χ^2/df	CFI	TLI	RMSEA	90% CI RMSEA
				406.420(266)	.000	1.528	.747	.715	.107	.084-.127
Emotional			.586							
3. Somatic	.755	.648								
8. Worries	.809	.795								
13.Unhappy	.763	.713								
16. Clingy	.843	.826								
24. Afraid	.878	.800								
Conduct			.432							
5. Tantrum	.891	.760								
7. Obedient	.434	.302								
12. Fights	.668	.679								
18. Lies	.747	.710								
22. Steals	.695	.726								
Hyperactivity			.441							
2. Restless	.645	.644								
10. Fidgety	.591	.689								
15. Distracted	.908	.670								
21. Reflects	.606	.648								

25. Attends	.687	.669	
Peer			.337
6. Loner	.435	.454	
11.Friend	.676	.767	
14. Popular	.426	.524	
19. Bullied	.890	.769	
23. Oldbest	.141	.153	
Pro social			.489
1. Considerate	.937	.730	
4. Shares	.430	.591	
9. Caring	.632	.722	
17. Kind	.937	.806	
20. Helpout	.702	.626	

Table 8 Comparison of CFA of Factor Models for the Self-Report SDQ-NSL and SDQ-NOR with Goodness of Fit Indices

Model	Version	$\chi^2(df)$	p	χ^2/df	CFI	TLI	RMSEA	90% CI RMSEA
SDQ-5-factor[1]	NSL	323.766(267)	.010	1.213	.801	.776	.073	.038-.100
	NOR	406.420(266)	<.001	1.528	.747	.715	.107	.084-.127
SDQ-1-factor[2]	NSL	218.116(170)	.008	1.283	.846	.827	.084	.046-.115
	NOR	286.070(170)	<.001	1.683	.768	.740	.122	.097-.146
SDQ-2nd-order[3]	NSL	333.091(273)	.008	1.220	.789	.769	.074	.041-.101
	NOR	423.994(271)	<.001	1.565	.724	.695	.111	.090-.131

[1]SDQ-5-factor model based on the five proposed subscales
[2]SDQ-1-factor model based on the four problems subscales included in the total score
[3]SDQ-2nd-order model based on the five subscales as well as a second-order total score for the four problem subscales

Table 9 Spearman's rho between Subscales and Total Score of the SDQ-NSL and SDQ-NOR Self-Report (N = 40)

	SDQ-NSL M (SD)	SDQ-NOR M (SD)	Spearman's rho**
Emotional Problems	4.40 (2.42)	4.15 (2.70)	.660
Conduct Problems*	2.35 (2.41)	1.55 (2.40)	.509
Hyperactivity	3.82 (1.90)	3.52 (1.81)	.538
Peer Problems	2.78 (1.73)	2.85 (1.76)	.599
Pro-social Behavior	8.18 (1.47)	8.15 (2.05)	.507
Total Difficulties	13.35 (6.28)	12.08 (6.83)	.668

*Significant difference on reported mean for Conduct Problems between the two self-report forms
**all Spearman's rho have $p < .001$

Table 10 Spearman Rank Correlations for the Self-and Parent report of the SDQ-NSL and SDQ-NOR.

	SDQ-NSL – parent SDQ-NOR (N = 26)	SDQ-NOR – parent SDQ-NOR (N = 30)
Emotional Problems	.521*	.400*
Conduct Problems	.043	.170
Hyperactivity	.318	.126
Peer Problems	.182	.351
Pro social Behavior	.026	-.029
Total Difficulties	.231	.269

* Correlations significant at $p < .05$

Table 11 Symptoms in the Normal, Sub-clinical and Clinical Range based on SDQ Self-Reports

Classification N (%)		SDQ-NOR		
SDQ-NSL	Normal	Sub-clinical	Clinical	Total
Normal	26 (65.0%)	1 (2.5%)	0 (0.0%)	27 (67.5%)
Sub-clinical	3 (7.5%)	1 (2.5%)	2 (5.0%)	6 (15.0%)
Clinical	2 (5.0%)	0 (0.0%)	5 (12.5%)	7 (17.5%)
Total	31 (77.5%)	2 (5.0%)	7 (17.5%)	40 (100%)

Appendix A

Detailed information on the CFA of the SDQ-NSL presented in Table 7

As can be seen in Table 7, the factor loading for item 11. “I have one good friend or more,” is negative. We analyzed the bivariate table for items 11 and 18 “I am often accused of lying and cheating.” Results are shown in Table 12.

<Table 12 here>

The empty cells for item 11 for “Not true” are due to the small sample size and cause the negative factor loading.

Table 12 Bivariate Table for Items 11 and 18 of the SDQ-NSL

		Item 18: I am often accused of lying and cheating		
Item 11: I have one good friend or more	Certainly true	Somewhat true	Not true	Total
	Certainly true	2	11	22	35
	Somewhat true	0	0	5	5
	Not true	0	0	0	0
Total		27	11	2	40

Appendix B

Table 13 Contingency table for preferred mode of communication in everyday life and preference for the SDQ-NSL or SDQ-NOR.

Communication in everyday life	Which version of the SDQ do you prefer?				
	NSL	Written	Combined	Don't know	Total
Spoken Norwegian	2	7	2	1	12
NSL	2	0	2	0	4
Bilingual	1	6	1	2	10
Total	5	13	5	3	26

NSL: Norwegian Sign Language

Fisher's exact test: $p = .196$

Table 14 Contingency table for sign language skills and preference for the SDQ-NSL or SDQ-NOR.

Sign language skills	Which version of the SDQ do you prefer?				
	NSL	Written	Combined	Don't know	Total
4	1	0	0	0	1
5	0	1	0	0	1
7	0	1	0	0	1
8	0	2	2	1	5
9	0	2	0	1	3
10	1	2	1	1	5
11	2	2	0	0	4
12	0	3	1	0	4
Total	4	13	4	3	24

NSL: Norwegian Sign Language
Sign language skills: Sum score of SUS and SPS, range 0 to 12;

Kruskal Wallis equality-of-populations rank test, chi-squared: p = .423

Table 15 Contingency table for spoken language skills and preference for the SDQ-NSL or SDQ-NOR.

Spoken language skills			Which version of the SDQ do you prefer?		
	NSL	Written	Combined	Don't know	Total
7	1	0	0	0	1
10	0	2	0	0	2
11	0	1	1	0	2
12	2	10	4	3	2
Total	3	13	5	3	19

NSL: Norwegian Sign Language
Spoken language skills: Sum score of CAP and SIR, range 0 to 12

Kruskal Wallis equality-of-populations rank test, chi-squared: p = .431