Article

Evaluation of a Novel Parent-Rated Scale for Selective Mutism

Assessment
2020, Vol. 27(5) 1007–1015

DOI: 10.1177/1073191118787328
journals.sagepub.com/home/asm
SAGE

Angelika Gensthaler[1], Julia Dieter[1], Susanne Raisig[2], Boris Hartmann[3], Marc Ligges[4], Michael Kaess[5,6], Christine M. Freitag[1], and Christina Schwenck[1,7]

Abstract

Assessment of selective mutism (SM) is hampered by the lack of diagnostic measures. The Frankfurt Scale of Selective Mutism was developed for kindergarteners, schoolchildren, and adolescents, including the diagnostic scale (DS) and the severity scale (SS). The objective of this study was to evaluate this novel, parent-rated questionnaire among individuals aged 3 to 18 years (n = 334) with SM, social phobia, internalizing disorders, and a control group. Item analysis resulted in high item-total correlations, and internal consistency in both scales was excellent with Cronbach's α = .90-.98. Exploratory factor analysis of the SS consistently yielded a one-factor solution. Mean sum scores of the DS differed significantly between the diagnostic groups, and the receiver operating characteristic analysis resulted in optimal cutoffs for distinguishing SM from all other groups with the area under the curves of 0.94-1.00. The SS sum scores correlated significantly with SM's clinician-rated symptom severity.

Keywords

selective mutism, measure development/validation, psychometric properties, reliability, validity

Despite speaking in other situations, children with selective mutism (SM) consistently fail to speak in specific situations where speaking is socially expected (American Psychiatric Association, 2013). SM is considered a rare childhood mental condition with prevalence rates of 0.7% to 2% (Bergman, Piacentini, & McCracken, 2002; Elizur & Perednik, 2003; Kumpulainen, Rasanen, Raaska, & Samppi, 1998). Furthermore, due to the long-standing lack of scientific evidence, the anxiety-based nature of this disorder was acknowledged only recently, resulting in the novel assignment of SM to the *Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition*'s (*DSM-5*) anxiety disorder section (American Psychiatric Association, 2013). Thus, few clinicians are familiar with its presentation or sufficiently trained in how to diagnose SM accurately via clinical interview and then initiate suitable therapy. Because of the general lack of knowledge about this debilitating disorder and of standardized, validated diagnostic measures, there is evidence that SM is often diagnosed late (Black & Uhde, 1995; Dummit et al., 1997; Schwartz, Freedy, & Sheridan, 2006), although its early diagnosis and treatment is essential to achieve remission: Older age is known to be associated with a worse outcome following pharmacological and psychotherapeutic interventions (Manassis, Oerbeck, & Overgaard, 2016; Oerbeck, Stein, Pripp, & Kristensen, 2015).

In addition to clinicians being inadequately trained, SM's diagnosis is compounded by its phenomenological overlap with social phobia (SP) (Muris & Ollendick, 2015), which reveals 97% to 100% comorbidity rates in children with SM (Gensthaler et al., 2016; Oerbeck, Stein, Wentzel-Larsen, Langsrud, & Kristensen, 2014; Veccio & Kearney, 2005). Differential diagnosis thus requires careful evaluation of the speaking habits to distinguish children with SM from those with SP alone. Moreover, even experts in this field have had difficulty developing specific and appropriate screening questions: The evaluation of a novel measure demonstrated that specificity of SM items was rather low concerning social anxiety symptoms in face validity checks by research experts and clinicians (Muris et al., 2017). The internal consistency of the resulting scales revealed in

[1]Goethe University, Frankfurt, Germany
[2]Humboldt University, Berlin, Germany
[3]Institut für Sprachtherapie, Cologne, Germany
[4]Jena University Hospital, Jena, Germany
[5]Ruprecht-Karls University, Heidelberg, Germany
[6]University Hospital of Child & Adolescent Psychiatry and Psychotherapy, Bern, Switzerland
[7]Justus-Liebig University, Gießen, Germany

Corresponding Author:
Angelika Gensthaler, Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt am Main, Deutschordenstraße 50, 60528 Frankfurt, Germany.
Email: angelika.gensthaler@kgu.de

clinical ($\alpha = .55$) and nonclinical samples ($\alpha = .65$) was also unsatisfactory.

Clinical evaluation is not yet supported by any validated diagnostic instrument, since no diagnostic cutoffs have been reported regarding the only questionnaire on SM that has been evaluated, namely the Selective Mutism Questionnaire (SMQ) (Bergman, Keller, Piacentini, & Bergman, 2008). Furthermore, the SMQ results' mean values did not distinguish SM from SP alone (Manassis et al., 2003). This is particularly problematic for clinicians and practitioners without specialized training in SM's assessment and treatment.

Additionally, even those clinicians who are sufficiently trained may have difficulty in efficiently assessing symptom severity, as evaluation of the individual speaking pattern covering all everyday social situations relevant to the child's psychosocial functioning is such a time-consuming procedure. Compounding the clinical diagnosis and estimation of symptom severity, mutism severity with the examining clinician does not correlate with parental or teachers' ratings of the mutism severity (Black & Uhde, 1995).

However, in addition to older age, both symptom severity (Oerbeck et al., 2015) and specific speaking patterns have been identified as predictors of poor outcome: In one study, the failure to speak with other children was identified as the sole predictor for a poor symptomatic outcome of SM (Steinhausen, Wachter, Laimböck, & Winkler, 2006). Another working group observed mute behavior within the core family to be the strongest predictor of a chronic course (Remschmidt, Poller, Herpertz-Dahlmann, Henninghausen, & Gutenbrunner, 2001). Reliably identifying subpopulations at risk may thus influence therapeutic decisions.

Comprehensive evaluation of the child's speaking pattern is also required for planning and monitoring behavioral interventions (e.g., exposure hierarchies, conversational visits, sliding-in procedures, play sessions with teachers or peers). Only a minority of the affected children and adolescents is completely mute, for example, at school (Kumpulainen et al., 1998). Most children with SM manage to communicate with certain teachers, children, or during breaks, for example (Bergman et al., 2002; Kumpulainen et al., 1998).

Further complicating this situation, mute behavior is modulated by various anxiety-provoking stimuli that are simultaneously inherent in social communication and that interact. McHolm, Cunningham, & Vanier (2005) groups them into three domains: person, place, and activity, hypothesizing that the communicative partner has the greatest impact on the failure to speak. Most children with SM find it more difficult to speak with unfamiliar persons; the degree of difficulty with familiar nonfamily members varies between that with strangers and close family members (Black & Uhde, 1995). The findings are inconclusive regarding the communicative partner's age: More children in the latter study remained mute toward adults than peers, whereas in the study by Kumpulainen and colleagues (1998), 45% of children remained mute toward all classmates at least in the school setting (vs. only 21% toward teachers).

Of all locations, it is school that most often elicits mute behavior—public places being less likely to do so—with the familiar home setting least likely (Black & Uhde, 1995). Also, certain contexts, such as speaking in front of others in a group or asking the teacher a question can trigger a mute reaction (Bergman et al., 2002). If verbal activities require little linguistic processing and if the child is confident about the answer (e.g., counting), little anxiety is invoked, whereas factual information or rehearsed speech requiring more verbal reasoning and ambiguity exacerbates anxiety (Johnson & Wintgens, 2001). Unplanned and social speech (greetings); having to provide alternatives, reasons, or opinions; and responses to ambiguous or difficult questions are extremely anxiety invoking. Furthermore, nonverbal communication may be contextually reduced or absent (Johnson & Wintgens, 2001).

To summarize, therapeutic interventions must be preceded by meticulous and time-consuming scrutiny of the child's speaking pattern. Clinicians cannot yet rely on a detailed validated scale for individual assessment. To date, the only evaluated questionnaire on SM is the SMQ, comprising 17 items with answers yielding a total score and subscale scores for the domains Home/Family, Public/Social, and School with internal consistencies of Cronbach's $\alpha = .84$ to $\alpha = .97$ (Bergman et al., 2008). In this study, SM children had higher mean scores in all domains than a mixed anxiety group. Letamendi et al. (2008) identified an internal consistency of Cronbach's $\alpha = .78$ for an abbreviated version (13 items) resulting from factor analysis. There is still no data on adolescents, as all studies to date only included children up to the age of 11 years. Moreover, younger age correlated with increased SM severity in the respective study (Bergman et al., 2008). Also, typically developing controls and a distinctive group of children with SP alone were not included in either investigation (Bergman et al., 2008; Letamendi et al., 2008). Although the SMQ has often been used in research, a revision of the measure as it is was recently proposed (Muris & Ollendick, 2015).

Given the need for a validated, age-adjusted measure of SM for both research and clinical practice, we developed the Frankfurt Scale of Selective Mutism (FSSM). The primary aim of this study was to report its relevant psychometric properties and to establish cutoff values in relation to the most important differential diagnoses for SM, SP alone, other internalizing disorders (INT), and typically developing controls. Second, we evaluated the measure for its capacity to assess symptom severity and individual speaking patterns in children with SM, which can serve as a starting point for an intervention and promote future research on SM.

We hypothesized that the measure would distinguish between SM and other internalizing conditions, particularly SP, and would reflect clinician-rated mutism severity independent of age. We further expected to replicate the three-factor model for contextual speaking dependent on location found by the SMQ.

Method

Participants

The $n = 334$ participants aged 3 to 18 years were part of a study on SM conducted in university departments of child and adolescent psychiatry, psychosomatics, and psychotherapy in Germany. The sample included participants with current SM ($n = 95$), SP ($n = 74$), INT ($n = 46$), and a typically developing control group (CG) ($n = 119$). A total of 89 (94%) children with ongoing SM also fulfilled the criteria of SP (Gensthaler et al., 2016). Subjects with a prior diagnosis of pervasive developmental disorder were excluded from the study. Intellectual delay (IQ < 70) was ruled out by a standardized test prior to participation if children did not attend regular classroom activities or were enrolled in special education programs. Mother-reported data on lifetime communication disorders of participants with SM and SP were gathered via diagnostic interviews and found to be equal (32%) in both groups. However, we did not collect data on the prevalence in CGs (Gensthaler et al., 2016). Other developmental abnormalities, particularly neuromuscular disorders impairing speech production, were ruled out by clinical evaluation and diagnostic interview. Recommendations for the SM diagnosis in bilingual individuals were respected (Toppelberg, Tabors, Coggins, Lum, & Burger, 2005). Of all the 95 participants with SM, $n = 28$ (29%) were raised bilingually, similar to the SP group (34%). For most of the bilingual SM participants ($n = 19$, 68%), German was their second language due to being immigrants, while only $n = 9$ (32%) bilingual SM children had at least one German native speaker as a parent. All but two bilingual participants with SM were born and raised in Germany. The INT group comprised participants with major depression (MD, $n = 14$), specific phobia ($n = 9$), obsessive–compulsive disorder (OCD, $n = 8$), generalized anxiety disorder (GAD, $n = 4$), separation anxiety disorder (SAD, $n = 4$), panic disorder ($n = 3$), anxiety disorder not otherwise specified ($n = 3$), and adjustment disorder ($n = 1$).

Our total sample was divided into three subsamples corresponding to the three age-adjusted versions of the FSSM: $n = 110$ preschoolers still attending kindergarten aged 3 to 7 years, including an SM ($n = 33$) and SP ($n = 17$) group and a CG ($n = 60$) of typically developing children; $n = 104$ school children aged 6 to 11 years, including SM ($n = 32$), SP ($n = 27$), INT ($n = 16$), and CG ($n = 29$); $n = 120$ adolescents aged 12 to 18 years, including SM, SP, INT, and CG, with $n = 30$ each. The preschool subsample lacked an INT group due to low referral rates.

The gender ratio was balanced within and between the diagnostic groups ($p = .180$-$.389$) in all the three age samples. The CG's socioeconomic status (SES) was significantly higher than that of all other groups in the preschool ($p = .012$) and adolescent ($p < .001$) sample but not in the school children sample (Gensthaler et al., 2016).

Mean age differed between groups in the preschool ($p = .001$) and school children ($p < .001$) samples. Post hoc analysis revealed that school-aged children in the INT group (10.6 years, $SD = 1.3$) were older than those in the SM group (9.1 years, $SD = 1.4$) and CG (8.8 years, $SD = 1.5$) but not the SP group (10.3 years, $SD = 1.1$). SP children were significantly younger than those in the CG. Earlier clinical referral of SM school children might also be responsible for the significantly lower mean age compared with school children with SP alone. In the preschool sample, the CG's mean age (4.5 years, $SD = 0.9$) according to post hoc analysis was lower compared with the SM (5.2 years, $SD = 0.9$) and SP groups (5.3 years, $SD = 0.8$).

Procedure

Most ($n = 85$) SM participants ($n = 95$) were recruited from university outpatient clinics. $N = 10$ were recruited from specialized therapeutic institutions for SM. Most participants with SP alone and INT were recruited from the same institutions. A community CG and some of the socially anxious preschoolers were recruited through contacts to kindergartens, schools, and a newspaper article. According to our study protocol, which was approved by the study center's local ethics committee, written informed consent was obtained from families. Families received no compensation for participation.

Prior to this study, all clinically referred participants were carefully diagnosed based on clinical evaluation, behavioral observation, parent and teacher reports, and psychological assessment. Within the framework of the present study, clinical diagnosis was confirmed and comorbid diagnoses identified in the SM and SP groups (Gensthaler et al., 2016) via a diagnostic interview. Fulfillment of the INT's *DSM-IV-TR* criteria was ensured by a review of medical records. To ensure correct grouping, a diagnostic interview was also conducted in case questionnaires indicated symptoms of SP or SM (INT and CG) or revealed clinical abnormalities (CG) (Gensthaler et al., 2016). Mothers completed the FSSM along with a set of other questionnaires.

Measures

Diagnostic Interview for Children and Adolescents (*Kinder-DIPS*): Current diagnostic status with regard to *DSM-IV-TR* (American Psychiatric Association, 2000) criteria was

evaluated using the structured *Kinder-DIPS* parent version, which has good retest and interrater reliability (Adornetto, In-Albon, & Schneider, 2008).

Additionally, symptom severity in SM children was estimated by the interviewer using the *Questionnaire on Social-Interactive Communication in Mutism* by Hartmann (2005), which is a clinician-rated, nonvalidated measure of SM based on 23 items (e.g., "Speaking with grandparents and other relatives") assessing communicative situations on a 3-point Likert-type scale (0 = *uninhibited communication*, 1 = *communication if asked to do so*, 2 = *selective/total muteness*) and yielding a sum score of 0 to 46. Since it is the only available German measure to date, it is widely used in clinical practice.

Item Development of the FSSM

The FSSM's item content and structure were developed by the first author. The questionnaire contains two independent sets of questions to be answered by a parent.

First, a diagnostic scale (DS) with 10 questions focusing on the presence of core characteristics of SM (e.g., *Is there a clear distinction between speaking behavior at home [rather talkative] and in public [avoiding the use of words or even mute]?*) was compiled. Items considering verbal (*Does your child not speak in certain situations and/or with certain persons, even though he/she is expected to?*) and nonverbal communication (*Is your child unable to shake or nod his/her head or point to something in certain situations, if asked to?*) as well as characteristic features (*Do his/her movements seem slow or frozen-like to you in certain situations?*) were included. Answers were scored as positive "Yes"—or negative "No"—answers, yielding a sum score equivalent to the number of positively answered items between 0 and 10.

As the DS neither evaluates SM's symptom severity nor identifies individual speaking patterns, a severity scale (SS) assessing the contextual failure to speak in a broad range of specific everyday situations (e.g., *Does your child speak with neighbors?*) was added. Questions refer to specific social interactions in kindergarten/school (e.g, *Does he/she speak with their father/mother in the kindergarten, even if others can hear them? Does he/she speak on the school playground? Does he/she read aloud in class?*), in public (*Does he/she speak with unfamiliar children when addressed on the playground, on holidays, or in a public swimming pool?*), and at home (*Does he/she speak at home with their parents' friends?*). SS items were grouped according to the three locations to increase applicability. Answers were given on a 5-point Likert-type scale (0 = *no problems*, 1= *with certain limitations [e.g., only avoiding the use of words or whispering or only when asked to do so]*, 2 = *occasionally*, 3= *hardly ever*, 4 = *not at all*), enabling us to calculate a total SS sum and subscale sums for the three domains.

In the third step, we revised the DS and SS to fulfill the demands of an age-specific measure. Adjusting and adding items (e.g., *Does your child speak with his/her babysitter? Is he/she able to have a short conversation out of politeness [small talk]?*) for specific age-groups resulted in three independent questionnaires of slightly different item content for children still attending kindergarten (FSSM 3-7), for school-aged children (FSSM 6-11), and for adolescents (FSSM 12-18). It is noteworthy that this also resulted in slightly different item numbers in the SS across FSSM versions.

Questionnaires were entered into data analysis if missing items did not exceed 10% per scale. This procedure resulted in the exclusion of three FSSM 3-7 and three FSSM 12-18. Missing data on the remaining questionnaires were imputed by the respective item means of the diagnostic group to allow for item analysis.

Statistical Analysis

The Statistical Package for the Social Sciences was used for data analyses. First, we tested the data distribution for the three FSSM versions and conducted item analyses. Internal consistency of the DS and the SS was tested by Cronbach's alpha.

Exploratory factor analyses of the SS in all versions were conducted separately to test for potential factor structure. Principal axis factor analysis (PFA) with eigenvalues greater than 1.0 and direct oblimin rotation were chosen because we expected to detect correlations among potential factors. As the eigenvalue-greater-than-1.0 rule tends to retain too many factors, scree tests were additionally conducted to confirm the number of factors to be retained (Costello & Osborne, 2005).

Prior to the analysis of mean sum scores of all FSSM scores and subscales, we assessed correlations between age and SES with scores of both scales via Pearson's coefficient and gender correlations with Spearman's rho scores. Analyses of covariances of the DS and multivariate analyses of covariances of the SS with age and SES as covariates were conducted to compare means of sum scores between groups. Post hoc analyses were calculated with *t* tests (Bonferroni).

To assess the FSSM's diagnostic accuracy, pairwise receiver operating characteristic (ROC) analyses on SM versus SP, INT, and CG, respectively, and SM versus a mixed SP–INT–CG group, most likely reflecting clinical reality, were calculated separately. Youden's index was used to identify optimal cutoff values of the DSs to differentiate groups. Correlations between the SS' sum scores and clinically rated severity were evaluated with Pearson's coefficient.

Results

Descriptive Statistics and Item Analysis

The DS and SS scores of the three FSSM versions were tested for normal distribution. Data demonstrated a

sufficiently normal distribution of scores according to the recommendations by Miles and Shevlin (2001). Skewness resulted in values for DS and SS in the slightly positive range (DS: .24-.60; SS: .82-.90); kurtosis showed values in the negative range (DS: −1.09 to −1.46; SS: −.48 to −.26).

FSSM results did not correlate with gender. We detected negative correlations between SES and DS and SS sum scores for the total preschool (DS: $r = -.24$, $p = .02$; SS: $r = -.28$, $p < .01$) and adolescent (DS: $r = -.25$, $p = .01$; SS: $r = -.35$, $p = .00$) samples. Furthermore, age correlated with higher DS ($r = .20$, $p = .04$) and SS ($r = .20$, $p = .04$) sum scores in our preschool sample. However, analyses of covariance for DS and SS revealed no significant influences by either SES (DS: preschool: $p = .71$, adolescent: $p = .35$; SS: preschool: $p = .44$, adolescent: $p = .20$) or age (DS: preschool: $p = .07$; SS: preschool: $p = .08$). We thus report analyses of variance results.

Item analysis revealed good item–total correlations (r^{it}) for the DS of all versions (FSSM 3-7: $r^{it} = .51$-$.80$, FSSM 6-11: $r^{it} = .49$-$.80$), and FSSM 12-18: $r^{it} = .34$-$.76$). Item analysis of the three SS showed a similarly high r^{it}, with a majority >.70. The items exhibited an r^{it} of .33 to .88 except for three items: item Home 1 (*Does your child speak with all close family members [mother, father, siblings]?,* $r^{it} = .26$) and Home 4 (*Does he/she speak about personal issues [feelings, wishes, conflicts, needs, decisions and experiences] at home?,* $r^{it} = .27$) of the FSSM 3-7 and again item Home 1 ($r^{it} = .15$) of the FSSM 6-11. Mutism within the core family is only prevalent in a small subset of SM children (Black & Uhde, 1995; Kumpulainen et al., 1998; Steinhausen & Juzi, 1996), which may also contribute to increased item difficulty and low item-total correlations in our mixed sample. It may indicate severe impairment and poor outcome (Remschmidt et al., 2001). In contrast to common practice, we did not exclude those items with $r^{it} < .30$, as those features should not be neglected when assessing SM youths.

Reliability

The DS's internal consistency in our entire sample ranged from α = .90 (FSSM 3-7, FSSM 12-18) to .92 (FSSM 6-11).

Internal consistency was α = .98 (Total SS), α = .95 (Kindergarten), α = .95 (Public), and α = .90 (Home) for the FSSM 3-7. The FSSM 6-11 yielded internal consistencies of α = .97 (Total SS), α = .96 (School), α = .96 (Public), and α = .79 (Home). Internal consistency was α = .98 (Total SS), α = .97 (School), α = .96 (Public), and α = .87 (Home) for the FSSM 12-18.

Internal consistencies of the SS in the target group of participants with SM were lower with α = .88 (Total SS), α = .84 (Kindergarten), α = .78 (Public), and α = .80 (Home) for the FSSM 3-7; α = .94 (Total SS), α = .93 (School), α = .90 (Public), and α = .73 (Home) for the FSSM 6-11; and α = .94 (Total SS), α = .90 (School), α = .87 (Public), and α = .84 (Home) for the FSSM 12-18. Interitem correlations of all SS items in the same diagnostic group (aiming to assess heterogeneous speaking patterns in SM) ranged from −.50 to .88 (mean .15) in the FSSM 3-7, −.34 to .86 (mean .26) in the FSSM 6-11; and −.36 to .86 (mean .28) in the FSSM 12-18, respectively.

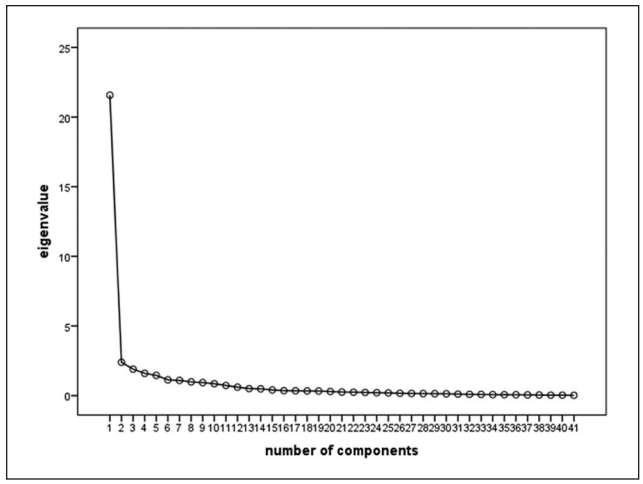

Figure 1. Scree test: Frankfurt Scale of Selective Mutism 3-7 Severity Scale.

Construct Validity of the Severity Scales

Unrestricted factor analysis of the items of the SSs of the three versions of the FSSM resulted in a high number of factors in each (FSSM 3-7: seven factors, FSSM 6-11: eight factors, and FSSM 12-18: six factors), which accounted for 70.7% (FSSM 3-7), 70.4% (FSSM 6-11), and 71.1% (FSSM 12-18) of the total variance. However, only the first factor seemed to be solid regarding item number, loadings, and absence of cross-loadings. None of the other factors fulfilled minimal quality criteria. Furthermore, as they were hard to interpret as distinct and meaningful factors, we carried out a scree test and subsequent PFA's with restricted factor numbers, as recommended by Costello and Osborne (2005) for such scenarios. The scree test clearly revealed a one-factor solution for the SSs of the three FSSM versions (see Figures 1-3). Subsequent PFA, with factor numbers restricted to two and one, respectively, identified a single-factor structure explaining 51.6% (FSSM 3-7), 48.5% (FSSM 6-11), and 55.2% (FSSM 12-18) of the total variance (see Supplementary Table S1, available in the online version of the article). Overall, our results deliver evidence of solid factor characteristics with a few, prominent exceptions: items Home 1 and 4, which already stood out from our item analysis due to low r^{it}, displayed low communalities or factor loadings or both across all three versions. Also

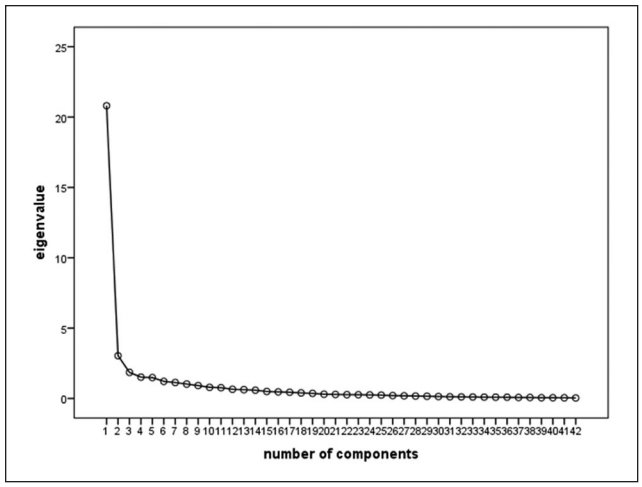

Figure 2. Scree test: Frankfurt Scale of Selective Mutism 6-11 Severity Scale.

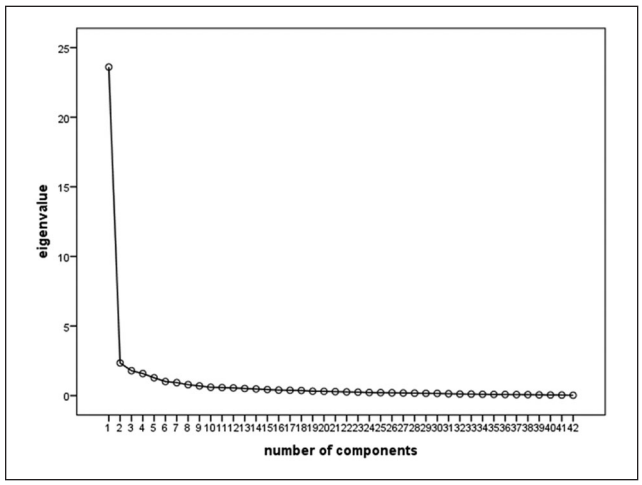

Figure 3. Scree test: Frankfurt Scale of Selective Mutism 12-18 Severity Scale.

items Home 7 (*Does he/she speak at home with his/her friends?*), School 4 (*Does your child play with a few select peers?*), and School 10 (*Does your child take part in gym class?*) of the FSSM 6-11 displayed low communality. In the FSSM 3-7, the corresponding items Kindergarten 4 (*Does your child join in sports activities?*) displayed low communality and Kindergarten 7 (*Does your child play with a few select peers?*) a low communality and factor loading.

Differential Validity

To test the DS for differential validity, sum scores were compared between the four diagnostic groups (analysis of covariances). Results are presented in Table 1. As expected, participants with SM scored considerably higher than all other groups, and sum scores differed significantly between groups. Post hoc analyses consistently revealed a significant pattern of sum scores of SM > SP > INT = CG (FSSM 3-7: SM > SP > CG) for all FSSM versions.

Furthermore, optimal cutoff values were evaluated based on ROC curve analyses. Results are shown in Table 2. In general, SM participants were readily distinguishable from SP, INT, and the CG. With regard to differentiating children with SM and SP, the three DSs of the FSSM exhibited satisfactory discriminating properties. Optimal cutoff values for the discrimination of SM and SP ranged from 6 (FSSM 12-18) to 7 (FSSM 3-7 and FSSM 6-11) with areas under the curve (AUC) = .94 to .99. Sensitivity was found to lie in the range of 84% for children attending kindergarten, 94% for school children, and 96% for adolescents. The specificity ranged from 94% for children in kindergarten to 93% for school children, to 72% for adolescents. Optimal cutoff values for the comparison between SM and INT and the CG, respectively, were considerably lower with consistent sensitivities of 100% and specificities between 94% and 100%. The differentiation of SM and a combined control group of all groups (SP + INT + CG) showed optimal cutoff values ranging from 6 (FSSM 3-7 and FSSM 12-18) to 7 (FSSM 6-11) with AUC = .97 to .99, sensitivities of 94% to 97%, and specificities of 90% to 95%.

Clinician-rated mutism severity in the SM groups based on the sum score of the measure by Hartmann (2005) significantly correlated with our results of the total SS scores of the three FSSM versions (FSSM 3-7: $r = .48$, $p < .01$, FSSM 6-11: $r = .72$, $p = .01$, FSSM 12-18: $r = .53$, $p = .01$), indicating good convergent validity.

Descriptive data of means and results of the between-group differences of the SS are displayed with Supplementary Table S2 (available in the online version of the article).

Discussion

Our findings reveal the FSSM's sound psychometric properties. Internal consistencies and item-total correlations of the DS and SS in our total sample were particularly high, both highlighting the underlying construct's strong coherence. Internal consistency markedly exceeded the SMQ's, which was assessed primarily in symptomatic children (Bergman et al., 2008; Letamendi et al., 2008). Given the lower Cronbach's alphas in the SSs of our group with SM alone, our results may be biased by deviant sample composition. Also, a higher item number, which clearly distinguishes both measures, is known to increase internal consistency (Cortina, 1993). Remarkably, the means of the SS's interitem correlations in the group of SM participants alone were below those reported concerning the 13-item version of the SMQ ($r = .32$-$.68$) (Letamendi et al., 2008), possibly indicating a convincing depiction of speaking patterns in the SS.

Table 1. FSSM Group Means of the Diagnostic Scale (DS).

	SM, mean (*SD*)	SP, mean (*SD*)	INT, mean (*SD*)	CG, mean (*SD*)	*df*	*F*	*p*	η^2	Post hoc, Bonferroni
FSSM 3-7	*n* = 31, 8.2 (1.5)	*n* = 17, 4.3 (1.5)	*n* = 0	*n* = 59, 1.3 (1.4)	2, 10	226.25	<.01	.81	SM > SP > CG
FSSM 6-11	*n* = 32, 8.7 (1.3)	*n* = 27, 3.7 (2.3)	*n* = 16, 1.6 (2.2)	*n* = 29, 1.2 (2.0)	3, 10	92.50	<.01	.74	SM > SP > INT = CG
FSSM 12-18	*n* = 28, 8.2 (1.5)	*n* = 29, 3.6 (2.5)	*n* = 30, 1.6 (2.0)	*n* = 30, 0.5 (0.8)	3, 11	100.10	<.01	.73	SM > SP > INT = CG

Note. FSSM = Frankfurt Scale of Selective Mutism; SM = selective mutism; SP = social phobia; INT = internalizing disorders; CG = control group.

Table 2. ROC Analyses of the Diagnostic Scales of the FSSM With Optimal Cutoff Values.

FSSM	Group comparison	Optimal Cutoff	Cutoff SE_{100}	Cutoff SP_{100}	ROC-AUC [95% CI]	Sensitivity[a]	Specificity[a]
FSSM 3-7	SM vs. SP	7	5	8	.97** [.92-1.00]	.84	.94
	SM vs. CG	5	5	6	1.00** [1.00-1.00]	1.00	.98
	SM vs. SP + CG	6	5	8	.99** [.98-1.00]	.97	.95
FSSM 6-11	SM vs. SP	7	6	8	.99** [.96-1.00]	.94	.93
	SM vs. INT	6	6	8	.99** [.97-1.00]	1.00	.94
	SM vs. CG	6	6	9	.99** [.97-1.00]	1.00	.97
	SM vs. SP + INT + CG	7	6	9	.99** [.97-1.00]	.94	.94
FSSM 12-18	SM vs. SP	6	5	9	.94** [.89-1.00]	.96	.72
	SM vs. INT	6	5	10	.98** [.93-1.00]	.96	.97
	SM vs. CG	4	4	4	1.00** [1.00-1.00]	1.00	1.00
	SM vs. SP + INT + CG	6	5	10	.97** [.95-1.00]	.96	.90

Note. FSSM = Frankfurt Scale of Selective Mutism; ROC = receiver operating characteristic; CI = confidence interval; SM = selective mutism; SP = social phobia; INT = internalizing disorders; CG = control group; SE_{100} = sensitivity of 100%; SP_{100} = specificity of 100%; AUC = area under the curve.
[a]At optimal cutoff.
**Significant at $p < .001$.

As hypothesized, differences in the DSs' mean values of diagnostic groups were highly significant with large effect sizes. Moreover, ROC-curve analysis resulted in applicable cutoffs with high specificities and sensitivities, also regarding the differentiation of children with SM from those with SP alone. Contrary to existing scales, which only assess the failure to speak in specific situations (Bergman et al., 2008; Manassis et al., 2003; Muris et al., 2017), the effective discrimination of participants with SM and SP alone may be explained by the focus of the DS's items on the SM's pathognomic feature—namely the consistently dichotomous speaking pattern. Diagnostic cutoffs and effect sizes of results of SMQ mean values of participants with SM and other anxiety disorders were not reported (Bergman et al., 2008), making the interpretation of our results in the context of existing literature difficult.

As expected, we found that SM's clinician-rated severity correlated positively with the total SS sum scores of all three versions of the FSSM. Due to the lack of a specific external criterion for SM's symptom severity in the studies evaluating the SMQ (Bergman et al., 2008; Letamendi et al., 2008), it is also difficult to compare our findings. However, in contrast to the SMQ's results (Bergman et al., 2008), we noted that younger age did not correlate with higher sum scores within the FSSM's three age samples, indicating developmentally appropriate item content. The wide range of interitem correlations (including negative correlations) within the SM group might indicate that individual, heterogeneous speaking behaviors in different social contexts are captured by the measure's items. In the future, the SS may thus serve in both clinical practice and research as a measure of severity and, on the item level, as an instrument to assess individual speaking patterns, promoting the research of their impact on course and outcome in particular and thereby facilitating therapeutic decisions.

Contrary to our assumptions, the SS's exploratory factor analysis yielded strong evidence of a single coherent factor of overall speaking patterns. This result contradicts previous studies on the SMQ (Bergman et al., 2008; Letamendi et al., 2008), which found three domains reflecting contextual speaking. Divergent findings may be due to several factors.

First, contrary to the FSSM, the SMQ's psychometric properties were examined in samples including either a majority of or exclusively target children with SM. Thus, the multidimensionality of contextual speaking patterns specific to SM is more likely to be detected than in our mixed sample, which included only a minority of participants with SM. Factor analysis in the group of SM alone was not done due to our small sample size.

Second, from a theoretical perspective, interaction among the three different anxiety-invoking stimuli in SM discussed above (inherent to each communicative situation) may play an important role in factor analysis, depending on the item content. Compared with the SMQ, the SS's items are considerably more specific about the speaking context, benefiting from the larger item number. For example, speaking habits with familiar peers or parents are evaluated across all three locations and in different contexts, since this information is the prerequisite for developing exposure hierarchies and choosing a therapeutic strategy (e.g., conversational visits vs. sliding-in procedures and play sessions with peers), its location, the verbal activity expected, and the key person involved (Johnson & Wintgens, 2001; McHolm et al., 2005). The more specific the item formulation concerning person, location, and verbal activity, the more overlap a specific feature will inevitably reveal (e.g., familiarity with the communicational partner) between items. While the accuracy of our assessment of the individual speaking patterns is improved, we are less likely to be able to extract dominant, distinct, and independent factors that underlie the situational failure to speak. A more specific item content and formulation may thus also account for divergent factor analytic findings.

Future users should keep in mind that SS subscales exclusively based on the location were retained preliminarily to maintain practical applicability and were not supported by psychometric analysis so far. As a result, our SS subscales' sum scores should be interpreted with caution until further analyses have been published.

This study's conclusions reveal a number of strengths and limitations. Above all, in spite of having enrolled a large overall sample, psychometric analysis was compromised by small sample sizes for the respective FSSM versions. As a result, we did not conduct factor analyses exclusively using data from symptomatic participants. Our results might thus be biased and comparison with existing scales confounded. Future evaluation of our measure should include the examination of its psychometric properties in a larger sample of target children. On the other hand, this study was conducted with clinically referred and thus preselected participants so that examiners were not blind to their diagnostic status. Replicating our study findings in population-based, nonreferred cohorts might rule out effects on results. The test-retest and inter-rater reliability of our measure must also be investigated. Our interpretation of the SS's convergent validity is limited by a clinician's rating based on an unevaluated scale. The lack of a valid external criterion, however, underscores the urgent need for our novel diagnostic tool. It will be up to future investigations to demonstrate whether the SS can be established as a measure of symptom change.

Our study suggests that the FSSM possesses solid psychometric properties regarding reliability and validity, particularly in the diagnostic accuracy of distinguishing children and adolescents with SM from those with SP alone. The parent-rated measure is efficient enough to serve as a valuable and easy-to-administer diagnostic questionnaire. Furthermore, it may promote research on SM in adolescents, for which no measure was available until now, and facilitate future evaluations of treatment outcomes. Our novel questionnaire may thus prove valuable for clinical practice and research alike.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

The online supplementary material is available at http://journals.sagepub.com/doi/suppl/10.1177/1073191118787328

References

Adornetto, C., In-Albon, T., & Schneider, S. (2008). Diagnostik im Kindes und Jugendalter anhand strukturierter Interviews: Anwendung und Durchführung des Kinder-Dips [Diagnostics in childhood and adolescence based on structured interviews: Application and implementation of the children's DIPS]. *Klinische Diagnostik und Evaluation*, *1*, 363-377.

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

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Bergman, R. L., Keller, M. L., Piacentini, J., & Bergman, A. J. (2008). The development and properties of the Selective Mutism Questionnaire. *Journal of Clinical Child & Adolescent Psychology*, *37*, 456-464. doi:10.1080/15374410801955805

Bergman, R. L., Piacentini, J., & McCracken, J. T. (2002). Prevalence and description of selective mutism in a school-based sample. *Journal of the American Academy of Child & Adolescent Psychiatry*, *41*, 938-946. doi:10.1097/00004583-200208000-00012

Black, B., & Uhde, T. W. (1995). Psychiatric characteristics of children with selective mutism: A pilot study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*, 847-855. doi:10.1097/00004583-199507000-00007

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*, 89-104. doi:10.1037/0021-9010.78.1.98

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation*, *10*(7). Retrieved from https://pareonline.net/pdf/v10n7.pdf

Dummit, E. S., Klein, R. G., Tancer, N. K., Asche, B., Martin, J., & Fairbanks, J. A. (1997). Systematic assessment of 50 children with selective mutism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 653-660. doi:10.1097/00004583-199705000-00016

Elizur, Y., & Perednik, R. (2003). Prevalence and description of selective mutism in immigrant and native families: a controlled study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*, 1451-1459.

Gensthaler, A, Maichrowitz, V, Kaess, M, Ligges, M, Freitag, C. M., & Schwenck, C. (2016). Selective mutism: The fraternal twin of childhood social phobia. *Psychopathology*, *49*, 95-107. doi:10.1159/000444882

Hartmann, B. (2005). *Evaluationsbogen für das sozialinteraktive Kommunikationsverhalten bei Mutismus* [Evaluation sheet for the socially interactive communication behavior in mutism]. Retrieved from http://boris-hartmann.de/sprachtherapie-boris-hartmann/files/evaluationsbogen3.pdf

Johnson, M., & Wintgens, A. (2001). *The selective mutism resource manual*. Brackley, England: Speechmark.

Kumpulainen, K., Rasanen, R., Raaska, H., & Samppi, V. (1998). Selective mutism among second-graders in an elementary school. *European Journal of Child & Adolescent Psychiatry*, *7*, 24-29. doi:10.1007/s007870050041

Letamendi, A. M., Chavira, D. A., Hitchcock, C. A., Roesch, S. C., Shipon-Blum, E., & Stein, M. B. (2008). Selective Mutism Questionnaire: Measurement structure and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 1197-1204. doi:10.1097/CHI.0b013e3181825a7b

Manassis, K., Fung, D., Tannock, R., Sloman, L., Fiksenbaum, L., & McInnes, A. (2003). Characterizing selective mutism: Is it more than social anxiety? *Depression and Anxiety*, *18*, 153-161. doi:10.1002/da.10125

Manassis, K., Oerbeck, B., & Overgaard, K. R. (2016). The use of medication in selective mutism: A systematic review. *European Journal of Child & Adolescent Psychiatry*, *25*, 571-578. doi:10.1007/s00787-015-0794-1

McHolm, A., Cunningham, C. E., & Vanier, M. K. (2005). *Helping your child with selective mutism*. Oakland, CA: New Harbinger.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers*. London, England: Sage.

Muris, P., & Ollendick, T. H. (2015). Children who are anxious in silence: A review on selective mutism, the new anxiety disorder in *DSM-5*. *Clinical Child and Family Psychology Review*, *18*, 151-169. doi:10.1007/s10567-015-0181-y

Muris, P., Simon, E., Lijphart, H., Bos, A., Hale, W., & Schmeitz, K. International Child and Adolescent Anxiety Assessment Expert Group (ICAAAEG). (2017). The Youth Anxiety Measure for *DSM-5* (YAM-5): Development and first psychometric evidence of a new scale for assessing anxiety disorders symptoms of children and adolescents. *Child Psychiatry & Human Development*, *48*, 1-17.

Oerbeck, B., Stein, M. B., Pripp, A. H., & Kristensen, H. (2015). Selective mutism: Follow-up study 1 year after end of treatment. *European Journal of Child & Adolescent Psychiatry*, *24*, 757-766. doi:10.1007/s00787-014-0620-1

Oerbeck, B., Stein, M. B., Wentzel-Larsen, T., Langsrud, Ø., & Kristensen, H. (2014). A randomized controlled trial of a home and school-based intervention for selective mutism: Defocused communication and behavioural techniques. *Child and Adolescent Mental Health*, *19*, 192-198. doi:10.1111/camh.12045

Remschmidt, H., Poller, M., Herpertz-Dahlmann, B., Henninghausen, K., & Gutenbrunner, C. (2001). A follow-up study of 45 patients with elective mutism. *European Archives of Psychiatry & Clinical Neuroscience*, *251*, 284-286. doi:10.1007/PL00007547

Schwartz, R. H., Freedy, A. S., & Sheridan, M. J. (2006). Selective mutism: Are primary care physicians missing the silence? *Clinical Pediatrics*, *45*, 43-48.

Steinhausen, H. C., & Juzi, C. (1996). Elective mutism: An analysis of 100 cases. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*, 606-614. doi:10.1097/00004583-199605000-00015

Steinhausen, H. C., Wachter, M., Laimböck, K., & Winkler Metzke, C. (2006). A long-term outcome study of selective mutism in childhood. *Journal of Child Psychology and Psychiatry*, *47*, 751-756. doi:10.1111/j.1469-7610.2005.01560.x

Toppelberg, C. O., Tabors, P., Coggins, A., Lum, K., & Burger, C. (2005). Differential diagnosis of selective mutism in bilingual children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *44*, 592-595. doi:10.1097/01.chi.0000157549.87078.f8

Veccio, J. L., & Kearney, C. A. (2005). Selective mutism in children: Comparison to youths with and without anxiety disorders. *Journal of Psychopathology and Behavioral Assessment*, *27*, 31-37. doi:10.1007/s10862-005-3263-1