

Die deutsche Selbstbeurteilungs- version des Strengths and Difficulties Questionnaire (SDQ-Deu-S)

Psychometrische Eigenschaften, Faktorenstruktur und Grenzwerte

Annette Lohbeck, Jan Schultheiß, Franz Petermann und Ulrike Petermann

Zusammenfassung. Die vorliegende Studie überprüft die Faktorenstruktur, psychometrischen Eigenschaften und Grenzwerte der deutschen Selbstbeurteilungsversion des Strengths and Difficulties Questionnaire (SDQ-Deu-S). Die Daten dieser Studie basieren auf einer Stichprobe von $N = 1\,501$ Kindern und Jugendlichen im Alter von 11 bis 16 Jahren. Die angenommene Fünf-Faktoren-Struktur des SDQ-Deu-S konnte anhand exploratorischer und konfirmatorischer Faktorenanalysen bestätigt werden. Die Reliabilitäten der fünf Skalen (Cronbachs Alpha $.55-.77$, Retest $.58-.67$) fielen jedoch gering aus. Lediglich die SDQ-Gesamtskala erreichte eine zufriedenstellende interne Konsistenz. Es werden alters- und geschlechtsspezifische Grenzwerte vorgeschlagen, die sich an die von Goodman (1997) empfohlenen Zielvorgaben von 80–10–10% für die Bestimmung von Grenzwerten in die drei Kategorien *unauffällig*, *grenzwertig* und *auffällig* anlehnen: Die Grenzwerte für die meisten SDQ-Subskalen liegen um etwa einen Punkt und für die SDQ-Gesamtskala um max. zwei Punkte niedriger als die Grenzwerte für die SDQ-Selbstberichtsversion von Goodman, Meltzer und Bailey (1998).

Schlüsselwörter. Verhaltensauffälligkeiten, psychometrische Eigenschaften, Faktorenstruktur, Grenzwerte, SDQ

The German Self-Report Version of the Strengths and Difficulties Questionnaire (SDQ-Deu-S): Psychometric Properties, Factor Structure, and Critical Values

Abstract. The present study examines factor structure, psychometric properties, and critical values of the German self-report version of the Strengths and Difficulties Questionnaire (SDQ-Deu-S). The data of this study are based on a sample of 1 501 children and adolescents aged 11 to 16 years. The assumed five-factor structure was confirmed by exploratory as well as confirmatory factor analyses. However, the reliability of the five scales (Cronbach's α $.55-.77$, retest $.58-.67$) was low. Only the Total Difficulties Scale revealed a satisfactory internal consistency. Gender- and age-specific critical values are suggested based on Goodman's (1997) target of having 80%, 10%, and 10% in the normal, borderline, and abnormal categories, respectively: The critical values for most of the SDQ subscales are one point lower and for the Total Difficulties Scale no more than two points lower than the recommended critical values for the British SDQ self-report version of Goodman, Meltzer, and Bailey (1998).

Keywords. Strengths and Difficulties Questionnaire, conduct problems, psychometric properties, factor structure, critical values, SDQ

Die Erfassung von Verhaltensauffälligkeiten und prosozialem Verhalten bei Kindern und Jugendlichen ist in der klinischen Diagnostik von großer Bedeutung. Neben der im deutschen Sprachraum weit verbreiteten *Child Behavior Checklist* (CBCL; Arbeitsgruppe Deutsche Child Behavior Checklist, 2012a, 2012b) gilt der *Strengths and Difficulties Questionnaire* (SDQ; Goodman, 1997) als ein international häufig eingesetztes Screening-Verfahren zur Erfassung verschiedener Verhaltensaspekte bei Kindern und Jugendlichen. Der SDQ kann sowohl die Stärken als auch die Verhaltensprobleme von Kindern und Jugendlichen im Alter von 2 bis 17 Jahren erfassen und liegt als Fragebogen in drei inhaltlich vergleichbaren Versionen für

Eltern, für Lehrkräfte/Erziehende sowie für Kinder/Jugendliche (ab 11 Jahre) vor. Im Internet ist der SDQ in über 50 Sprachen mit einer Anleitung zur Auswertung kostenfrei erhältlich (Strengths and Difficulties Questionnaire, 2015a). Die deutsche Selbstbeurteilungsversion des SDQ nach Goodman (2005), die auf der Website abrufbar ist und auch in der vorliegenden Arbeit angewendet wurde, soll in dieser Arbeit als SDQ-Deu-S bezeichnet werden. Der Fragebogen setzt sich aus 25 Items zusammen, die sich mit jeweils fünf Items auf die fünf Skalen *Emotionale Probleme* (EP), *Externalisierende Verhaltensauffälligkeiten* (VA), *Hyperaktivität/Aufmerksamkeitsprobleme* (HYP), *Probleme mit Gleichaltrigen* (PG)

und *Prosoziales Verhalten* (PRO) verteilen. Die vier Problemskalen EP, VA, HYP und PG können dabei als Gesamtproblemwert aufsummiert werden. Sowohl für die fünf Subskalen als auch für den Gesamtproblemwert liegt eine von Goodman (1997) eingeteilte Zuordnung zu auffälligen, grenzwertigen und unauffälligen Kategorien vor. Die Grenzwerte wurden anhand der jeweiligen Rohwertverteilung der britischen Normierungsstichprobe bestimmt, wobei etwa 80 % der Werte als unauffällig und jeweils 10 % als grenzwertig und auffällig eingestuft wurden.

Die Zusammenstellung dieser fünf Skalen konnte bereits in einer Vielzahl von Studien faktorenanalytisch belegt werden (z.B. Becker, Hagenberg, Roessner, Woerner & Rothenberger, 2004; vgl. Kóbor, Takás & Urbán, 2013; Koglin, Barquero, Mayer, Scheithauer & Petermann, 2007). Zahlreiche nationale und internationale Arbeiten berichten seit über 15 Jahren über die psychometrischen Eigenschaften des SDQ (z.B. Giannakopoulos et al., 2013; Gómez-Beneyto et al., 2013; Ruchkin, Kopusov, Vermeiren & Schwab-Stone, 2012). Die meisten Befunde beziehen sich allerdings vorrangig auf die SDQ-Elternversion und auf die SDQ-Lehrkräfteversion, nur für wenige der verschiedenen übersetzten SDQ-Versionen liegen auch Normen vor. Eine deutsche Normierung wurde sowohl für die SDQ-Elternversion (Woerner et al., 2002) und für die Lehrkräfteversion bzw. Erzieherinnen-/Erzieherversion (Koglin et al., 2007) als auch für die SDQ-Selbstbeurteilungsversion (Altendorfer-Kling, Ardelt-Gattinger & Thun-Hohenstein, 2007) vorgenommen. Die Normen für die deutsche SDQ-Selbstbeurteilungsversion basieren jedoch lediglich auf Daten einer österreichischen Feldstichprobe von 2529 Kindern und Jugendlichen aus Hauptschulen und einer Handelsakademie in den Bundesländern Salzburg und Steiermark. Die vorliegende Arbeit möchte deshalb an diese Studie anknüpfen und die psychometrischen Eigenschaften sowie die Faktorenstruktur der deutschen SDQ-Selbstbeurteilungsversion ausschließlich an einer deutschen Stichprobe überprüfen, die sowohl Haupt- und Realschülerinnen bzw. Haupt- und Realschüler als auch Gymnasiasten berücksichtigt. Darüber hinaus soll anhand der Rohwertverteilung der Gesamtstichprobe alters- und geschlechtsspezifische Grenzwerte bestimmt werden, die sich an die von Goodman (1997) festgelegten alters- und geschlechtsübergreifenden Zielvorgaben von 80–10–10 % zur Bestimmung von Grenzwerten in die drei Bereiche *unauffällig*, *grenzwertig* und *auffällig* annähern sollen.

Wenngleich der SDQ als ein gut geprüftes Instrument zur Erfassung von Problemverhalten und prosozialem Verhalten gilt, liegen zum Teil sehr heterogene Ergebnisse zu den psychometrischen Eigenschaften des SDQ vor. Wenig konsistent sind vor allem die Befunde zur Reliabilität der fünf Skalen: In einer Pilotstudie von Goodman, Meltzer und Bailey (1998) ließen sich z. B. mit der SDQ-Selbstbeurteilungsversion sehr stark variierende inter-

ne Konsistenzen für die fünf Skalen EP ($\alpha = .75$), VA ($\alpha = .72$), HYP ($\alpha = .69$), PRO ($\alpha = .65$), PG ($\alpha = .61$) und für den SDQ-Gesamtproblemwert ($\alpha = .82$) feststellen. Auch in anderen Studien (Capron, Théron & Duyme, 2007; Muris, Meesters, Eijkelenboom & Vincken, 2004; vgl. Riso et al., 2010) finden sich mehrheitlich eher geringe Cronbachs α -Werte für die fünf Skalen EP ($.63 \leq \alpha \leq .66$), VA ($.47 \leq \alpha \leq .60$), HYP ($.66 \leq \alpha \leq .67$), PRO ($.59 \leq \alpha \leq .66$), PG ($.39 \leq \alpha \leq .46$) und für den SDQ-Gesamtproblemwert ($.70 \leq \alpha \leq .80$). Dies verdeutlicht nicht zuletzt die Studie von Essau et al. (2012), in der die SDQ-Selbstbeurteilungsversion bei 2418 Jugendlichen zwischen 12 und 17 Jahren in fünf verschiedenen Ländern (Deutschland, Zypern, England, Schweden, Italien) angewendet wurde: Für die deutsche SDQ-Selbstbeurteilungsversion schwankten die Cronbachs α -Werte in dieser Studie von .45 bis .74, für die zyprische SDQ-Version von .07 bis .83, für die englische SDQ-Version von .26 bis .71, für die schwedische SDQ-Version von .02 bis .68 und für die italienische SDQ-Version lagen diese zwischen .22 und .68. Ähnliche Befunde zeigen auch Riso et al. (2010), die in einer italienischen Stichprobe von 1 349 Grundschulkindern im Alter von 8 bis 10 Jahren Cronbachs α -Werte von .27 bis .67 für die fünf SDQ-Skalen fanden. Die Retest-Reliabilitäten der fünf Skalen scheinen in den meisten Studien weitgehend zufriedenstellend zu sein (z.B. Lai et al. 2010; Liu et al., 2013), auch wenn sie generell im niedrigen Bereich liegen: In einer Studie von Giannakopoulos et al. (2013) werden z. B. Intraklassenkorrelationen (ICC) von .65 bis .89 berichtet, und auch bei Liu et al. (2013) bewegen sich diese zwischen .67 und .89 eher im niedrigen Bereich. Darüber hinaus weist eine Reihe von Studien mehrheitlich auf geringe bis mittlere Korrelationen zwischen Eltern-, Lehrkräfte- und Selbstbeurteilungen hin (Borg, Kaukonen, Salmelin, Joukamaa & Tamminen, 2012; Mieloo et al., 2014; Stone, Otten, Engels, Vermulst & Janssens, 2010).

Verschiedene Studien zur faktoriellen Validität des SDQ haben bereits anhand *exploratorischer Faktorenanalysen* (EFA) die von Goodman (1997) mitgeteilte Fünf-Faktoren-Struktur konsistent belegt (Altendorfer-Kling et al., 2007; Becker et al., 2006; Capron et al., 2007; Giannakopoulos et al., 2013; Goodman, 2001; Koglin et al., 2007; Mansbach-Kleinfeld, Apter, Farbstein, Levine & Ponizovsky, 2010). Ausgehend von der postulierten Fünf-Faktoren-Struktur des SDQ wurden vermehrt auch *konfirmatorische Faktorenanalysen* (CFA) zur Überprüfung der SDQ-Faktoren-Struktur durchgeführt, aus denen jedoch sehr unterschiedliche Ergebnisse resultierten (Goodman, Lamping & Ploubidis, 2010; Van Roy, Veenstra & Clench-Aas, 2008; vgl. Kóbor et al., 2013): Während einige Studien eine Fünf-Faktoren-Struktur bestätigen (He, Burstein, Schmitz & Merikangas, 2013; Liu et al., 2013; Richter, Sagatun, Heyerdahl, Oppedal & Røysamb, 2011; Van Roy et al., 2008; Yao et al., 2009), legen andere

Studien eine Drei-Faktoren-Struktur nahe (Dicky & Blumberg, 2004; Koskelainen, Sourander & Vauras, 2001). Die drei Faktoren bilden danach (a) *Internalisierende Verhaltensprobleme* mit den beiden SDQ-Skalen EP und PG, (b) *Externalisierende Verhaltensprobleme* mit den beiden SDQ-Skalen VA und HYP und schließlich (c) die SDQ-Skala PRO. Diese Drei-Faktoren-Struktur hat sich auch in den neueren Studien von Riso et al. (2010) und von Goodman et al. (2010) als angemessener erwiesen als die Fünf-Faktoren-Struktur. Bei Vergleich der fünf verschiedenen SDQ-Selbstbeurteilungsversionen aus Deutschland, Zypern, England, Schweden und Italien zeigten sich in der Studie von Essau et al. (2012) dagegen sehr uneinheitliche Ergebnisse: Während das Drei-Faktoren-Modell in dem SDQ aus Zypern den besten Modellfit erreichte, wies das Fünf-Faktoren-Modell im deutschen SDQ die beste Modellanpassung auf. In den übrigen drei Ländern fielen die Fit-Indizes jedoch für beide Modelle wenig zufriedenstellend aus und in der Gesamtstichprobe ($N = 2418$) ergaben sich nahezu identische Fit-Indizes für diese beiden Modelle.

Aufgrund der bislang einzigen Normierung der deutschen SDQ-Selbstbeurteilungsversion an der österreichischen Feldstichprobe von Altendorfer-Kling et al. (2007) und der mangelhaften Befundlage zu den psychometrischen Eigenschaften des SDQ-Deu-S ist es ein Anliegen der vorliegenden Studie, die Faktorenstruktur und psychometrischen Eigenschaften des SDQ-Deu-S ausschließlich an einer deutschen Stichprobe zu überprüfen. Berücksichtigt werden sollen dabei folgende Punkte:

1. die psychometrischen Kennwerte,
2. die faktorielle Validität durch Überprüfung des Fünf-Faktoren-Modells sowie eines Drei-Faktoren-Modells anhand exploratorischer und konfirmatorischer Faktorenanalysen und
3. die Retest-Reliabilitäten.

Zudem sollen anhand der Rohwertverteilung der vorliegenden Stichprobe alters- und geschlechtsspezifische Grenzwerte bestimmt werden, die sich an die von Goodman (1997) festgelegten Zielvorgaben von 80–10–10% (jeweils 10% für die Kategorie *grenzwertig* und *auffällig*) möglichst gut annähern. Zusätzlich werden Alters-, Geschlechts- und Schicht- bzw. Schulformeffekte untersucht, da diese in allen drei bisherigen Studien zum deutschen SDQ bereits hinlänglich bekannt sind (vgl. Altendorfer-Kling et al., 2007; Koglin et al., 2007; Wörner et al., 2002). Angesichts der kritischen und widersprüchlichen nationalen und internationalen Befunde zu den psychometrischen Eigenschaften der SDQ-Selbstbeurteilungsversion sollen jedoch für die vorliegende Studie keine soliden Annahmen zur Reliabilität des SDQ-Deu-S formuliert werden. Den Befunden von Essau et al. (2012) entsprechend wird dennoch erwartet, dass sich für den vorliegenden SDQ-Deu-S eine Fünf-Faktoren-Faktorenstruktur in den exploratorischen und konfirmatorischen Faktorenanalysen zeigt.

Methode

Die vorliegende Studie basiert auf Daten einer umfassenden Schulstichprobe von $N = 1501$ Kindern und Jugendlichen im Alter von 11 bis 16 Jahren ($M = 13.85$, $SD = 1.53$) aus Niedersachsen und Bremen. Es handelt sich demnach um eine sehr altersheterogene und regionale Schulstichprobe. Diese Stichprobe setzt sich aus 789 Jungen (53%) und 712 Mädchen (47%) zusammen, die aus zwei Teilstichproben stammen: 845 Daten (56%, Jungen: $n = 442$; Mädchen: $n = 403$, $M = 14.38$, $SD = 1.01$) wurden aus einer Evaluationsstudie zum JobFit-Training an Schulen verwendet (Laakmann, Schultheiß, Petermann & Petermann, 2013), in der 273 Daten von Kindern und Jugendlichen ausgewertet wurden. Für die vorliegende Studie konnte ein größerer Datensatz von 845 Fällen bereitgestellt werden, da das JobFit-Training in der Zwischenzeit noch weiter evaluiert wurde. Die übrigen 656 Daten (44%, Jungen: $n = 347$, Mädchen: $n = 309$, $M = 13.15$, $SD = 1.79$) stammen von Schülerinnen und Schülern verschiedener Schulformen (Gymnasium, Oberschule, Haupt- und Realschule) aus einer Validierungsstudie zu einem Schüler selbstbeurteilungsverfahren (Lohbeck, Nitkowski, Petermann & Petermann, 2014). Diese Studie umfasste 1267 Schülerinnen und Schüler zwischen 9 und 19 Jahren. In die vorliegende Analysestichprobe gingen nur solche Schülerinnen und Schüler aus diesen beiden Studien ein, die zum Zeitpunkt der Erhebung zwischen 11 und 16 Jahren alt waren und bei denen der Anteil an fehlenden Werten (*missing values*) für die einzelnen SDQ-Items unter 10% lag. 1141 Schülerinnen und Schüler kamen aus Sekundarschulen und 360 waren Gymnasiasten. Tabelle 1 schlüsselt die genaue Alters- und Geschlechterverteilung für die Gesamtstichprobe getrennt nach Schulformen (Gymnasium, Sekundarschulen) auf.

Zur Berechnung von Retest-Reliabilitäten konnten die Daten von $N = 911$ Kindern und Jugendlichen im Alter von 11 bis 16 Jahren ($M = 14.04$, $SD = 1.07$) ausgewertet werden, die 6 Wochen nach der ersten Erhebung erhoben wurden. Diese Retest-Stichprobe umfasst 465 Jungen (51%) und 446 Mädchen (49%).

Alle Datenerhebungen fanden sowohl für die Hauptals auch für die Retest-Untersuchung in Schulklassen unter standardisierter Anleitung von geschulten Testleitenden statt. Die Schülerinnen und Schüler füllten den Fragebogen eigenständig in einer regulären Unterrichtsstunde im Klassenverband unter Anwesenheit eines Testleitenden und einer Klassenlehrkraft aus. Das Einverständnis der Eltern wurde eine Woche vor der Datenerhebung in den Schulen von allen beteiligten Kindern und Jugendlichen eingeholt. Eine Genehmigung der Landes Schulbehörde Niedersachsen für die Durchführung der Datenerhebungen an den Schulen lag vor. Die Teilnahme an der Befragung war freiwillig und anonym.

Tabelle 1. Alters- und Geschlechterverteilung für die Gesamtstichprobe getrennt nach Schulformen

Schulform			Alter						Gesamtsumme
			11	12	13	14	15	16	
Sekundarschulen (Oberschule, Hauptschule, Realschule)	Jungen	Anzahl	41	56	122	126	143	120	608
		%	7	9	20	21	23	20	100
	Mädchen	Anzahl	24	55	114	137	104	99	533
		%	4	10	21	26	20	19	100
	Gesamtsumme	Anzahl	65	111	236	263	247	219	1141
		%	6	10	21	23	21	19	100
Gymnasium	Jungen	Anzahl	37	23	42	35	13	31	181
		%	21	13	23	19	7	17	100
	Mädchen	Anzahl	31	28	54	27	13	26	179
		%	17	16	30	15	7	15	100
	Gesamtsumme	Anzahl	68	51	96	62	26	57	360
		%	19	14	27	17	7	16	100
Alle Schulformen	Jungen	Anzahl	78	79	164	161	156	151	789
		%	10	10	21	20	20	19	100
	Mädchen	Anzahl	55	83	168	164	117	125	712
		%	8	12	24	23	16	17	100
	Gesamtsumme	Anzahl	133	162	332	325	273	276	1501
		%	9	11	22	22	18	18	100

Messinstrument

Der eingangs beschriebene SDQ beinhaltet 25 Items und fünf Skalen, die sowohl die Verhaltensauffälligkeiten als auch das prosoziale Verhalten von Kindern und Jugendlichen zwischen 2 und 17 Jahren erfassen. In der vorliegenden Studie wurde die deutschsprachige SDQ-Selbstbeurteilungsversion für Kinder und Jugendliche angewendet. Jeweils fünf Items bildeten eine Skala:

1. *Emotionale Probleme* (Beispielitem: „Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen.“),
2. *Externalisierende Verhaltensauffälligkeiten* (Beispielitem: „Ich werde leicht wütend; ich verliere oft meine Beherrschung.“),
3. *Hyperaktivitäts- und Aufmerksamkeitsprobleme* (Beispielitem: „Ich lasse mich leicht ablenken; ich finde es schwer, mich zu konzentrieren.“),
4. *Probleme mit Gleichaltrigen* (Beispielitem: „Ich bin meistens alleine; ich beschäftige mich lieber mit mir selbst.“) und
5. *Prosoziales Verhalten* (Beispielitem: „Ich teile normalerweise mit Anderen.“).

Die fünf Skalen wurden durch Aufsummierung der jeweiligen fünf Items berechnet, wobei fünf Items der Problemskalen für die Auswertung umkodiert werden mussten, da diese positiv formuliert sind. Die vier Problemskalen wurden zu einem Gesamtproblemwert verrechnet. Die Beantwortung erfolgte auf einer dreistufigen

Likert-Skala (0 = nicht zutreffend, 1 = teilweise zutreffend, 2 = eindeutig zutreffend).

Datenauswertung

Die faktorielle Validität des SDQ-Deu-S wurde wie in den Studien von Koglin et al. (2007) und von Woerner et al. (2002) zunächst anhand exploratorischer Hauptkomponentenanalysen mit einer Varimax-Rotation und einer forcierten Fünf-Faktoren-Lösung für die Gesamtstichprobe überprüft. Zusätzlich wurden konfirmatorische Faktorenanalysen mit der *maximum likelihood*-Methode (Programm: AMOS 21; Arbuckle, 2012) berechnet, in denen zwei Modelle miteinander verglichen wurden: (a) ein Drei-Faktoren-Modell mit den drei korrelierten Faktoren *Externalisierende Verhaltensprobleme*, *Internalisierende Verhaltensprobleme* und *Prosoziales Verhalten* und (b) das angestrebte Fünf-Faktoren-Modell mit den fünf korrelierten Faktoren *Emotionale Probleme* (EP), *Verhaltensauffälligkeiten* (VA), *Hyperaktivität/Aufmerksamkeitsprobleme* (HYP), *Probleme mit Gleichaltrigen* (PG) und *Prosoziales Verhalten* (PRO). Die Beurteilung der Modellgüte erfolgte anhand der klassischen Fit-Indizes (*comparative fit index* [CFI], *tucker lewis index* [TLI], *root mean square error of approximation* [RMSEA], ergänzt um χ^2 und *df*). Ein guter Modellfit liegt bei folgenden Schwellenwerten vor (vgl. Weiber & Mühlhaus, 2014): $\chi^2/df \leq 3$, $CFI \geq 0.9$, $TLI \geq 0.9$, $RMSEA \leq 0.05$). Zur Überprüfung der psychometrischen Eigenschaften des SDQ-Deu-S wurden Skalenmittelwerte, -streuungen und

Itemtrennschärfen berechnet. Die internen Konsistenzen der fünf SDQ-Skalen wurden nach Cronbachs α bestimmt. Retest-Reliabilitäten, Interkorrelationen und Korrelationen mit dem Alter, Geschlecht und der Schulform wurden über Spearman-Rangkorrelationen untersucht. Zur Überprüfung von Alters-, Geschlechts- und Schicht- bzw. Schulformunterschieden in den Skalenmittelwerten dienten *Mann-Whitney U-Tests*. Aufgrund der sehr altersheterogenen Stichprobe wurden vor den Analysen zwei altershomogenere Gruppen gebildet (11;0 bis 13;11 Jahre und 14;0 bis 16;11 Jahre). Schichteffekte wurden über die Schulformzugehörigkeit der Schülerinnen und Schüler analysiert, da keine soziodemografischen Angaben wie z. B. der Migrationshintergrund oder die Deutschsprachkenntnisse der Schülerinnen und Schüler vorlagen. Haupt- und Realschulen wurden mit Oberschulen zu „Sekundarschulen“ zusammengefasst, da in Oberschulen sowohl Haupt- als auch Realschülerinnen bzw. Haupt- und Realschüler unterrichtet werden und diese Schulform die Haupt- und Realschule in Niedersachsen zukünftig ablösen soll. *Missing values* in den Items wurden durch den Mittelwert der zugehörigen Skala ersetzt. Nach Goodman (1997) wurde ein Hochrechnen der fünf Items erst dann durchgeführt, wenn mindestens drei Items der jeweiligen Skala vorlagen. Datensätze mit mehr als 10% *missing values* wurden von der Analyse ausgeschlossen. 124 Datensätze wurden wegen zu vieler Einzel-*missings* (> 2 je Subskala) aus den beiden Teilstichproben bereits vor den Analysen eliminiert (23 Datensätze aus der Evaluationsstudie und 101 Datensätze aus der Validierungsstudie). In die Berechnungen gingen somit ausschließlich komplett ausgefüllte Fragebögen und Fragebögen mit maximal zwei ersetzten *missings* ein. Angaben zur Dropout-Rate bzw. zur Beteiligung der Schülerinnen und Schüler in den einzelnen Schulklassen wurden nicht dokumentiert. Zur Bestimmung geeigneter Grenzwerte für die vorliegende Stichprobe in die drei Kategorien *unauffällig*, *grenzwertig* und *auffällig* wurden die von Goodman (1997) festgelegten Zielvorgaben von 80–10–10% als Orientierung herangezogen, wonach jeweils 10% auf die Kategorien *auffällig* und *grenzwertig* fallen. Um zunächst die Passung der Grenzwerte, die für die Auswertung des SDQ auf der Website zu finden sind (Strengths and Difficulties Questionnaire, 2015b), für die vorliegende Stichprobe zu beurteilen, wurden die Daten als erstes anhand der dort vorgeschlagenen Grenzwerte ausgewertet. Diese Grenzwerte basieren auf der Studie von Goodman et al. (1998), die allerdings nur an einer Stichprobe von 83 Kindern bestimmt wurden. Da bei dieser Auswertung nach britischen Grenzwerten die kombinierten Prozentanteile der Kategorien *grenzwertig* und *auffällig* sowohl insgesamt als auch bei Jungen und Mädchen in den Altersgruppen 11–13 und 14–16 für mehrere Skalen des SDQ-Deu-S weit unter 20% lagen und damit nicht den Zielvorgaben von 80–10–10% entsprachen, wurden anschließend für mehrere Skalen des SDQ-Deu-S neue Grenzwerte bestimmt. Angestrebt wurden dabei so-

wohl alters- als auch geschlechtsspezifische Grenzwerte, da eine sehr altersheterogene Stichprobe untersucht wurde und von Geschlechtseffekten ausgegangen werden konnte. Auf schulformspezifische Grenzwerte wurde hingegen verzichtet, da der SDQ als schulformübergreifendes Instrument eingesetzt wird und sich dadurch eine bessere Vergleichbarkeit mit den meisten bisherigen Studien zum SDQ gewährleisten lässt. Das heißt: In der vorliegenden Studie wurde insbesondere darauf geachtet, dass in die Kategorie *auffällig* maximal 10,5% fallen. Für den Bereich *grenzwertig* wurden im Vergleich zu den britischen Grenzwerten für die meisten SDQ-Skalen etwas geringere Grenzwerte mit zwei Punkten gewählt, damit in den Subskalen nicht weniger als 20% für die Kategorien *grenzwertig* oder *auffällig* erreicht werden. Für den Fall, dass der *wahre* Grenzwert zwischen zwei möglichen Rohwertpunkten lag, wurden diejenigen Grenzwerte herangezogen, die sicherstellen, dass in den Subskalen nicht weniger als 19% in die Kategorien *grenzwertig* und *auffällig* fallen. Die Absicht dieser Studie war demnach eine Fokussierung auf eine hohe Sensitivität, verbunden mit dem Ziel, dass *auffällige* Werte nur von wenigen Kindern erzielt werden und in der Kategorie *grenzwertig* kein Kind bzw. kein Jugendlicher mit möglicherweise problematischen Werten übersehen wird.

Ergebnisse

Faktorenstruktur

Die postulierte Fünf-Faktoren-Struktur des SDQ lässt sich in einer varimax-rotierten Hauptkomponentenanalyse mit einer voreingestellten fünffaktoriellen Extraktionsmethode weitgehend reproduzieren. Die meisten SDQ-Items laden auf dem Faktor am höchsten, der auch der jeweiligen SDQ-Skala entspricht. Lediglich die beiden Items 21 und 25 der Skala *Hyperaktivität* und das Item 7 der Skala *Verhaltensauffälligkeiten* weisen eine negative Nebenladung von über -0.40 auf. Auch das negativ gepolte und umkodierte Item 14 der Skala *Probleme mit Gleichaltrigen* ist mit einer negativen Nebenladung von -0.37 auf dem nicht korrespondierenden Faktor 5 als kritisch anzusehen. Dennoch lassen sich nach dem Kaiser-Eigenwertkriterium genau fünf Faktoren mit einem Eigenwert von > 1 extrahieren (Eigenwertverlauf: 4.01, 2.63, 1.79, 1.41, 1.23). Die Varianzaufklärung für diese fünf Faktoren beträgt etwa 44% (1. Faktor: 16%, 2. Faktor: 10%, 3. Faktor: 7%, 4. Faktor: 6%, 5. Faktor: 5%). Ebenso weisen die *Measure of Sample Adequacy*-Koeffizienten der 25 SDQ-Items von $.74$ bis $.91$ und der *Kaiser-Meyer-Olkin*-Koeffizient mit $.82$ auf die Angemessenheit der Itemzuordnung hin. Dies unterstreicht zudem der *Bartlett*-Test, der signifikant ausfällt ($\chi^2 = 6213.87$, $df = 300$, $p < .001$). Die Kommunalitäten der 25 Items variieren allerdings stark zwischen $.25$ und $.66$ im moderaten Bereich.

Tabelle 2. Interne Konsistenzen (Cronbachs α), Retest-Korrelationen (r_{tt}) und part-whole korrigierte Itemtrennschärfen ($r_{i(t-1)}$) der fünf SDQ-Skalen

Skala	Cronbachs α			Retest: r_{tt}			Trennschärfe: $r_{i(t-1)}$		
	Gesamt $N = 1501$	Jungen $n = 789$	Mädchen $n = 712$	Gesamt $N = 911$	Jungen $n = 465$	Mädchen $n = 446$	Gesamt $N = 1501$	Jungen $n = 789$	Mädchen $n = 712$
EP	.71	.69	.69	.63	.54	.65	.36–.53	.33–.50	.34–.51
VA	.55	.56	.51	.59	.55	.61	.18–.36	.20–.37	.15–.37
HYP	.66	.68	.64	.67	.66	.68	.28–.50	.30–.53	.25–.51
PG	.56	.56	.56	.58	.52	.64	.27–.37	.24–.35	.29–.38
PRO	.63	.61	.60	.62	.61	.57	.29–.47	.26–.44	.26–.47
GP	.77	.79	.76	.66	.60	.73	.18–.46	.19–.48	.12–.44

Anmerkungen: N = Größe der Gesamtstichprobe, n = Größe der Teilstichprobe Jungen bzw. Mädchen, SDQ = Strengths and Difficulties Questionnaire, EP = Emotionale Probleme, VA = Verhaltensauffälligkeiten, HYP = Hyperaktivität, PG = Probleme mit Gleichaltrigen, PRO = Prosoziales Verhalten, GP = Gesamtwert. Alle Retest-Korrelationen sind auf dem Niveau von $p < .001$ signifikant. Das Retest-Intervall beträgt 6 Wochen. Die Trennschärfen beziehen sich auf die Analysen der Hauptuntersuchung.

Die Überprüfung der Modellstruktur in den konfirmatorischen Faktorenanalysen fällt dagegen weniger überzeugend aus: Weder das postulierte Fünf-Faktoren-Modell noch das in einigen Studien nahegelegte Drei-Faktoren-Modell ist mit den Daten hinreichend kompatibel. Das Fünf-Faktoren-Modell zeigt dennoch eine deutlich bessere Anpassung an die Daten ($\chi^2/df = 5.12$, $df = 265$, CFI = .82, TLI = .79, RMSEA = .05, standardisierte Regressionskoeffizienten: min. $\beta = .30$, max. $\beta = .66$) als das Drei-Faktoren-Modell mit den drei korrelierten Faktoren *Externalisierende Verhaltensprobleme*, *Internalisierende Verhaltensprobleme* und *Prosoziales Verhalten* ($\chi^2/df = 7.00$, $df = 272$, CFI = .73, TLI = .70, RMSEA = .06, standardisierte Regressionskoeffizienten: min. $\beta = .22$, max. $\beta = .64$).

Skaleninterkorrelationen

Zwischen den vier SDQ-Problemskalen lassen sich sehr stark variierende positive Korrelationen feststellen: Die manifesten auf den Rohwerten basierten Interkorrelationen der vier Skalen schwanken von .13 bis .45 und die geschätzten latenten Faktoreninterkorrelationen der konfirmatorischen Faktorenanalyse für das Fünf-Faktoren-Modell von .23 bis .65. Die Korrelationen zwischen den vier SDQ-Problemskalen und der SDQ-Skala *Prosoziales Verhalten* erreichen auf manifester Ebene Werte von -.31 bis .09 und auf latenter Ebene Werte von -.45 bis .18.

Psychometrische Skalenkennwerte

Die internen Konsistenzen der fünf SDQ-Skalen sind sowohl in der Gesamtstichprobe als auch bei Jungen und Mädchen niedrig ausgeprägt: In der Gesamtstichprobe liegen Cronbachs α -Werte von .55 bis .71 vor, bei Jungen von .56 bis .69 und bei Mädchen variieren diese von .51

bis .69. Für die SDQ-Gesamtproblemskala lassen sich etwas höhere Cronbachs α -Werte von .77 in der Gesamtstichprobe, von .79 bei Jungen und von .76 bei Mädchen ermitteln. Auch die Retest-Reliabilitäten der fünf Skalen bewegen sich eher im niedrigen Bereich: Sie betragen in der Gesamtstichprobe für die fünf Subskalen zwischen .58 und .67, bei Jungen zwischen .52 und .66 und bei Mädchen liegen diese zwischen .57 und .68. Die SDQ-Gesamtproblemskala erreicht in der Gesamtstichprobe einen Retest-Korrelationskoeffizienten von .66, bei Jungen einen Koeffizienten von .60 und bei Mädchen beträgt dieser .73. Ebenso sind die part-whole korrigierten Itemtrennschärfen bei Jungen und Mädchen sowie in der Gesamtstichprobe eher im geringen Bereich einzuordnen; teilweise liegen Itemtrennschärfen von $< .30$ vor (vgl. Tabelle 2).

Mittelwerte und Effekte von Alter, Geschlecht und Schulform

Die Mittelwerte der Skalen liegen generell in niedriger Größenordnung (Maximalwert der Subskalen: 10, Maximalwert der Gesamtproblemskala: 40). Die *Mann-Whitney U*-Tests und die Korrelationsanalysen mit den Variablen Geschlecht (Jungen = 0, Mädchen = 1), Alter und Schulform (Sekundarschulen = 0, Gymnasium = 1) deuten auf einige bedeutsame Geschlechts- und Schulforneffekte hin. Tabelle 3 zeigt die Ergebnisse dieser Analysen.

Während Jungen signifikant höhere Werte auf den Skalen VA und PG erzielen, weisen Mädchen substantiell höhere Werte auf den Skalen EP, PRO und GP auf. Bedeutsame Alterseffekte lassen sich dagegen für keine der SDQ-Skalen statistisch stützen, jedoch signifikante Schulforneffekte: Bis auf die Skala EP erreichen Gymnasiasten auf allen SDQ-Skalen deutlich niedrigere Werte als Schülerinnen und Schüler anderer Sekundarschulen (vgl. Tabelle 3).

Tabelle 3. Mittelwerte (Standardabweichungen) und Mittelwertunterschiede (Mann-Whitney U-Tests) sowie Korrelationen (Spearman-Rangkorrelationen) mit Geschlecht, Alter und Schultyp für die SDQ-Deu-S-Skalen

	Geschlecht			Altersgruppe		Schultyp		Korrelationen mit		
	Gesamt N = 1501	Jungen n = 789	Mädchen n = 712	11–13 n = 627	14–16 n = 874	OS/HS/RS n = 1141	GYM n = 360	Geschlecht	Alter	Schultyp
EP	2.73 (2.22)	2.14*** (1.99)	3.38 (2.27)	2.72 (2.22)	2.74 (2.22)	2.78 (2.26)	2.58 (2.06)	.29***	.00	-.03
VA	1.74 (1.54)	1.93*** (1.64)	1.52 (1.41)	1.72 (1.51)	1.75 (1.57)	1.85 (1.62)	1.39*** (1.24)	-.12***	-.00	-.11***
HYP	3.46 (2.08)	3.46 (2.14)	3.46 (2.00)	3.50 (2.02)	3.44 (2.12)	3.61 (2.09)	3.02*** (1.98)	.00	-.02	-.12***
PG	2.45 (1.70)	2.54* (1.73)	2.35 (1.66)	2.39 (1.79)	2.48 (1.63)	2.59 (1.72)	1.99*** (1.53)	-.06*	.05	-.15***
PRO	7.36 (1.83)	6.96*** (1.88)	7.81 (1.67)	7.40 (1.81)	7.33 (1.85)	7.27 (1.89)	7.65** (1.60)	.23***	-.02	-.08**
GP	10.37 (5.14)	10.07** (5.28)	10.71 (4.96)	10.34 (5.08)	10.40 (5.18)	10.82 (5.24)	8.98*** (4.57)	.07**	.00	-.15***

Anmerkungen: SDQ-Deu-S bezeichnet die deutsche Selbstbeurteilungsversion des Strengths and Difficulties Questionnaire nach Goodman (2005), die in dieser Studie ausschließlich an einer Schultstichprobe (S) überprüft wurde. N = Größe der Gesamtstichprobe, n = Größe der Teilstichproben, EP = Emotionale Probleme, VA = Verhaltensauffälligkeiten, HYP = Hyperaktivität, PG = Probleme mit Gleichaltrigen, PRO = Prosoziales Verhalten, GP = Gesamtproblemwert. OS = Oberschule, HS = Hauptschule, RS = Realschule. GYM = Gymnasium. ***p < .001, **p < .01, *p < .05.

Grenzwerte und Prozentanteile der Schülerinnen und Schüler in den Kategorien *grenzwertig* und *auffällig*

Bei Anwendung der von Goodman et al. (1998) empfohlenen alters- und geschlechtsunspezifischen Grenzwerte fällt in der vorliegenden Stichprobe auf, dass die Prozentanteile der Kategorien *grenzwertig* und *auffällig* für die meisten SDQ-Skalen nicht den Zielvorgaben von 80–10–10% entsprechen. Zum Teil liegen die kombinierten Prozentanteile für diese beiden Kategorien sogar unter 10%. Die bei Anwendung der von Goodman et al. (1998) empfohlenen Grenzwerte berechneten Prozentanteile für die Kategorien *grenzwertig* und *auffällig* in dieser Stichprobe können bei der Erstautorin angefordert werden.

Die in dieser Studie festgelegten alters- und geschlechtsspezifischen Grenzwerte, die anhand der Rohwertverteilung der vorliegenden Gesamtstichprobe bestimmt wurden und sich an die von Goodman (1997) empfohlenen Zielvorgaben von 80–10–10% mit den drei Kategorien *unauffällig*, *grenzwertig* und *auffällig* annähern sollten, stellt Tabelle 4 dar.

Im Vergleich zu den von Goodman et al. (1998) vorgeschlagenen alters- und geschlechtsunspezifischen Grenzwerten werden für die vorliegende Stichprobe etwas geringere Grenzwerte empfohlen. Der kombinierte Prozentanteil der Schülerinnen und Schüler für die Kategorien *grenzwertig* und *auffällig* fällt nach den in dieser Studie vorgeschlagenen Grenzwerten daher deutlich höher aus als bei Anwendung der britischen Grenzwerte von

Goodman et al. (1998). In der vorliegenden Gesamtstichprobe liegen für alle SDQ-Skalen in den beiden Altersgruppen mehr als 20% der Kinder und Jugendliche in den Kategorien *grenzwertig* und *auffällig*. Der höchste kombinierte Prozentanteil dieser beiden Kategorien ergibt sich in der jüngeren Altersgruppe mit 31% für die Skala HYP und in der älteren Altersgruppe für die Skala PRO mit 33%. Bei Jungen sind die Grenzwerte für die Kategorie *grenzwertig* in beiden Altersgruppen abgesehen von den Skalen PG und PRO um etwa einen Punkt und in der Kategorie *auffällig* bis auf die Skalen VA, HYP, PG und PRO um einen Punkt bzw. zwei Punkte (bei der Skala GP in der Altersgruppe 14–16) niedriger als die britischen Grenzwerte von Goodman et al. (1998). Am höchsten fällt der kombinierte Prozentanteil für die beiden kritischen Bereiche in der vorliegenden Stichprobe sowohl bei Jungen der jüngeren Altersgruppe (34%) als auch bei Jungen der älteren Altersgruppe (31%) für die SDQ-Skala HYP aus. Auch bei Mädchen sind die Grenzwerte in beiden Altersgruppen gegenüber den britischen Grenzwerten für die meisten SDQ-Skalen um etwa einen Punkt bzw. zwei Punkte (Skala GP) niedriger. Der höchste kombinierte Prozentanteil ergibt sich bei Mädchen der jüngeren Altersgruppe für die SDQ-Skala EP (30%) und bei Mädchen der älteren Altersgruppe für die SDQ-Skala HYP (33%). Eine anschauliche grafisch beschreibende Darstellung der Skalenmittelwerte sowie der SDQ-auffälligen und -grenzwertigen Prozentanteile je Altersgruppe und Geschlecht sowohl nach britischen als auch nach den hier vorgeschlagenen Grenzwerten zeigen die Abbildungen im Elektronischen Supplement 1 dieser Studie.

Table 4. Grenzwerte für die SDQ-Deu-S-Skalen und Prozentanteil der Schülerinnen und Schüler mit Werten in den Kategorien *grenzwertig* und *auffällig* in der vorliegenden Stichprobe

Altersgruppe	Gesamt (N = 627)					Jungen (n = 321)					Mädchen (n = 306)				
	G	G%	A	A%	G%+A%	G	G%	A	A%	G%+A%	G	G%	A	A%	G%+A%
11;0–13;11	5–6	16	7–10	7	23	4–5	17	6–10	9	26	5–6	20	7–10	10	30
	3–4	20	5–10	6	26	3–4	23	5–10	8	31	3	12	4–10	8	20
	5–6	24	7–10	7	31	5–6	25	7–10	9	34	5–6	24	7–10	5	29
	4–5	18	6–10	5	23	4–5	21	6–10	5	26	4–5	14	6–10	5	19
	5–6	24	0–4	6	30	5	15	0–4	9	24	6	14	0–5	8	22
	15–17	10	18–40	10	20	15–18	12	19–40	9	21	15–17	10	18–40	9	19
14;0–16;11	Gesamt (N = 874)					Jungen (n = 468)					Mädchen (n = 406)				
	G	G%	A	A%	G%+A%	G	G%	A	A%	G%+A%	G	G%	A	A%	G%+A%
	5–6	14	7–10	7	21	4–5	16	6–10	6	22	5–6	19	7–10	10	29
	3–4	17	5–10	8	25	3–4	20	5–10	9	29	3	10	4–10	9	19
	5–6	24	7–10	8	32	5–6	23	7–10	8	31	5–6	24	7–10	9	33
	4–5	21	6–10	5	26	4–5	24	6–10	5	29	4	13	5–10	10	23
	5–6	26	0–4	7	33	5	17	0–4	9	26	6	13	0–5	10	23
	15–17	11	18–40	10	21	15–17	9	18–40	9	18	15–17	9	18–40	11	20

Anmerkungen: SDQ-Deu-S bezeichnet die deutsche Selbstbeurteilungsversion des Strengths and Difficulties Questionnaire nach Goodman (2005), die in dieser Studie ausschließlich an einer Schultstichprobe (S) überprüft wurde. N = Größe der Gesamtstichprobe, n = Größe der Teilstichprobe Jungen bzw. Mädchen, G = Grenzwerte für die Kategorie *grenzwertig*, G% = Prozentanteil in der Kategorie *grenzwertig*, A = Grenzwerte für die Kategorie *auffällig*, A% = Prozentanteil in der Kategorie *auffällig*. Bei den Prozentanteilen handelt es sich um gerundete Werte. EP = Emotionale Probleme, VA = Verhaltensauffälligkeiten, HYP = Hyperaktivität, PG = Probleme mit Gleichaltrigen, PRO = Prosoziales Verhalten, GP = Gesamtproblemwert.

Diskussion

In der vorliegenden Studie wurden die Faktorenstruktur, die psychometrischen Eigenschaften und die Grenzwerte der deutschen Selbstbeurteilungsversion des *Strengths and Difficulties Questionnaire* (SDQ-Deu-S) an einer deutschen Schulstichprobe überprüft. Die zentralen Befunde dieser Studie lassen sich wie folgt zusammenfassen:

Die angenommene dimensionale Struktur des SDQ mit den fünf Subskalen *Emotionale Probleme*, *Verhaltensauffälligkeiten*, *Hyperaktivität/Aufmerksamkeitsprobleme*, *Probleme mit Gleichaltrigen* und *Prosoziales Verhalten* ließ sich in einer varimax-rotierten Hauptkomponentenanalyse für eine forcierte Fünf-Faktoren-Lösung weitgehend replizieren. Dennoch stellten sich bei einigen Items der fünf SDQ-Subskalen (Items 7, 21 und 25) auch substantielle höhere Ladungen auf anderen Faktoren heraus als auf dem erwarteten korrespondierenden Faktor, was zugleich darauf hindeutet, dass diese Items weniger gut geeignet sind. Auch die Modellprüfungen der konfirmatorischen Faktorenanalysen fielen weniger überzeugend aus: Das angestrebte Fünf-Faktoren-Modell wies zwar eine deutlich bessere Anpassung an die Daten auf als das in einigen Studien (z. B. Dicky & Blumberg, 2004; Koskelainen et al., 2001) nahegelegte Drei-Faktoren-Modell mit den drei Faktoren *Internalisierende Verhaltensprobleme*, *Externalisierende Probleme* und *Prosoziales Verhalten*. Die Fit-Indizes für das Fünf-Faktoren-Modell können jedoch in dieser Überprüfung nicht als ausreichend angesehen werden. Eine mögliche Erklärung für dieses nicht zufriedenstellende Ergebnis könnten eventuell die umkodierte Items sein, was auch schon einige Studien (z. B. Essau et al., 2012; Palmieri & Smith, 2007) nahelegen, denen zufolge die umkodierte Items die Faktorenstruktur verändern.

Als kritisch zu betrachten wären in der vorliegenden Studie zudem die psychometrischen Eigenschaften des SDQ-Deu-S: Die internen Konsistenzen nach Cronbachs α fielen sowohl für die fünf Subskalen als auch für die SDQ-Gesamtproblemskala generell niedriger aus als in der deutschen SDQ-Elternversion oder in der SDQ-Version für Erzieherinnen und Erzieher (Koglin et al., 2007; Woerner et al., 2002), auch wenn sie mit anderen Studien zum deutschen SDQ-Selbstberichtsverfahren annähernd übereinstimmen (Altendorfer-Kling et al., 2007; Essau et al., 2012). Ebenso lagen in der vorliegenden Studie die Itemtrennschärfen eher im niedrigen Bereich, teilweise fanden sich auch Itemtrennschärfen mit einem Wert von unter .30. In Einklang mit den internationalen Analysen lässt sich daraus letztlich nur den Schluss ziehen, dass der SDQ-Deu-S auf Subskalenebene eher weniger reliabel interpretierbar ist und eventuell eine Überarbeitung dieses Fragebogens mit einer Erweiterung der Items zu überlegen wäre. Dies gilt es jedoch zugleich insofern zu relativieren, als die Höhe der internen Konsistenz und der korrigierten Itemtrennschärfen auch sehr stark von der

Anzahl der Items und von der Enge des mit ihnen gemessenen Merkmals abhängig ist. Wenn lediglich fünf Items einen relativ breiten Merkmalsbereich erfassen, ist es schließlich wenig sinnvoll, diesbezüglich die gleichen Maßstäbe anzulegen wie bei einem Leistungstest, der z. B. mit 30 Items eine spezifische Fähigkeit erfassen möchte. Beim SDQ stellen die fünf Items einer Subskala zwar miteinander verwandte, aber eben doch sehr unterschiedliche Facetten des im Skalennamen angegebenen Merkmalsbereichs dar, was insbesondere für ein Screening-Verfahren in diesem Kontext sehr wünschenswert ist. In der Absicht, einen im Vergleich zur CBCL kürzeren und ökonomischeren Fragebogen zu entwickeln, der sich als vergleichbar valide erweist, mussten die fünf einzelnen SDQ-Items somit unterschiedliche Teilaspekte ansprechen. Ein Cronbachs α -Wert von .90 und durchweg hohe Trennschärfen hätten lediglich Redundanz angezeigt.

Die Mittelwerte der SDQ-Deu-S-Skalen lagen in vergleichbarer Größenordnung wie die der britischen SDQ-Selbstbeurteilungsversion oder die der anderen übersetzten Versionen, die im Internet auf der Website des SDQ zu finden sind (Strengths and Difficulties Questionnaire, 2015a); auch in der österreichischen Stichprobe bei Altendorfer-Kling et al. (2007) fielen sie nahezu identisch aus. Die stark variierenden Interkorrelationen können dabei als Beleg dafür gesehen werden, dass die fünf Subskalen im SDQ-Deu-S eindeutig voneinander abgrenzbar sind, was eine differenzierte Problemanalyse ermöglicht. Die positive Korrelation zwischen den Skalen *Emotionale Probleme* und *Prosoziales Verhalten*, die sich in der vorliegenden Studie ebenso abbildete, ist allerdings erklärungsbedürftig. Plausibel begründen lässt sich diese aber auch damit, dass Schülerinnen und Schüler, die emotionale Probleme kennen bzw. bei sich selbst gut wahrnehmen können, sich auch gut in die Lage von anderen Gleichaltrigen hineinversetzen dürften, wenn diese verletzt, krank oder traurig sind, sodass sie sich gerade deswegen auch prosozial gegenüber diesen Gleichaltrigen verhalten. Bei genauerer Betrachtung der einzelnen Itemformulierungen fällt zudem auf, dass sich für die Skala *Prosoziales Verhalten* auch mehrere Items auf emotionale Aspekte beziehen, wie z. B. das Item 1 „Ich versuche, nett zu anderen Menschen zu sein, ihre Gefühle sind mir wichtig.“ oder das Item 9 „Ich bin hilfsbereit, wenn andere verletzt, krank oder traurig sind.“. Demzufolge erklärt sich und ist gut nachvollziehbar, dass z. B. die Items der Skala *Emotionale Probleme* wie vor allem das Item 8 („Ich mache mir häufig Sorgen.“) oder das Item 13 („Ich bin oft unglücklich oder niedergeschlagen; ich muss häufig weinen.“) sehr gut zu den oben genannten Items 1 und 9 der Skala *Prosoziales Verhalten* passen und positiv miteinander korrelieren. Ein ähnliches Ergebnis berichten beispielsweise auch Lohbeck et al. (2014) in einer Studie, in der sich zwischen der SDQ-Skala *Emotionale Probleme* und der Skala *Einfühlungsvermögen* eine signifikant positive Korrelation von $r = .19$ ergab.

Für mehrere SDQ-Skalen ließen sich darüber hinaus signifikante Geschlechts- und Schulformeffekte feststellen, jedoch keine bedeutsamen Alterseffekte: Während Mädchen auf den SDQ-Skalen *Emotionale Probleme*, *Prosoziales Verhalten* und auf der SDQ-Gesamtpblemskala höhere Werte verzeichneten als Jungen, erzielten letztere deutlich höhere Werte auf den SDQ-Skalen *Verhaltensauffälligkeiten* und *Probleme mit Gleichaltrigen*. Besonders vorteilhaft scheint dabei die Zugehörigkeit zu einem Gymnasium zu sein, da Gymnasiasten bis auf die SDQ-Skala *Emotionale Probleme* auf sämtlichen SDQ-Skalen signifikant niedrigere Werte erreichten als Schülerinnen und Schüler aus anderen Sekundarschulen (Oberschule, Haupt- und Realschule). Dieser Schulformeffekt ist insofern gut nachvollziehbar, als naheliegend davon ausgegangen werden kann, dass Gymnasiasten in der Regel einen höheren sozioökonomischen Status besitzen und weniger Verhaltensprobleme zeigen. Zu berücksichtigen ist jedoch, dass die Mehrzahl der in dieser Studie untersuchten Schülerinnen und Schüler aus Sekundarschulen stammten ($n = 1\,141$).

Die Ergebnisse zu den in dieser Studie empfohlenen alters- und geschlechtsspezifischen Grenzwerten, die sich an die von Goodman (1997) angegebenen Zielvorgaben von 80–10–10% in die drei Bereiche *unauffällig*, *grenzwertig* und *auffällig* möglichst gut annähern sollten und anhand der Rohwertverteilung der Schülerinnen und Schüler in der vorliegenden Stichprobe bestimmt wurden, zeigten für alle SDQ-Skalen deutlich, dass sehr viele Schülerinnen und Schülern in die Kategorien *grenzwertig* und *auffällig* fallen, besonders für die Skala *Hyperaktivität*. Dies begründet sich offensichtlich damit, dass sich die Grenzwerte für die Kategorie *grenzwertig* in dieser Stichprobe weniger klar eingrenzen ließen und teilweise aus zwei Punkten gebildet wurden. Ein Beispiel: In der Skala *Hyperaktivität* sind in der vorliegenden Stichprobe bei den 14;0- bis 16;11-jährigen Mädchen 33% grenzwertig oder auffällig, da die Kategorie *grenzwertig* durch die beiden Rohwertpunkte 5 und 6 gebildet wurde. Beschränkt man die grenzwertige Kategorie jedoch auf den Rohwert 6, wären nur 16% entweder grenzwertig oder auffällig, da 17% dieser Gruppe den Rohwert 5 erreichen. Der Anteil grenzwertiger und auffälliger Kinder und Jugendlicher fiel in dieser Studie deshalb deutlich höher aus als in der Studie von Goodman et al. (1998). Eine Verringerung der Grenzwerte um einen Punkt bzw. max. zwei Punkte für die SDQ-Gesamtpblemskala gegenüber den britischen Grenzwerten von Goodman et al. (1998) kann für die vorliegende Stichprobe dennoch durchaus als angemessen angesehen werden, da bei Anwendung der britischen Grenzwerte weder für die SDQ-Gesamtpblemskala noch für die fünf SDQ-Subskalen die von Goodman (1997) festgelegten Zielvorgaben von 80–10–10% mit den drei Kategorien *unauffällig*, *grenzwertig* und *auffällig* erreicht werden. Anstatt der geforderten 20% grenzwertiger bzw. auffälliger Kinder ließen sich in den

Altersgruppen 11–13 und 14–16 bei Anwendung der britischen Grenzwerte sehr stark variierende kombinierte Prozentanteile zwischen 6% und 28% für die SDQ-Skalen feststellen. Bei Jungen zeigten sich z.B. in beiden Altersgruppen auf der Skala *Emotionale Probleme* und bei Mädchen auf den Skalen *Verhaltensauffälligkeiten* und *Prosoziales Verhalten* weniger als 10% grenzwertige und auffällige Werte. Dies würde in der Praxis allerdings bedeuten, dass Mädchen mit einem grenzwertigen prosozialem Verhalten oder einem auffälligen Verhalten möglicherweise durch den SDQ-Deu-S mit den britischen Normwerten „übersehen“ werden, genauso wie Jungen mit grenzwertigen emotionalen Problemen. Da die kombinierten Werte von *grenzwertig* und *auffällig* in dieser Stichprobe sogar noch unter 10% lagen, kann davon ausgegangen werden, dass bei Anwendung der britischen Grenzwerte von Goodman et al. (1998) auch einige als auffällig zu bezeichnende Kinder und Jugendliche vom SDQ-Deu-S als unauffällig eingestuft werden.

Um mit einem Screening-Instrument möglichst keine problematischen Werte zu übersehen, dürften die in dieser Studie gewählten geringeren Grenzwerte für die Kategorie *grenzwertig* insofern sinnvoll sein. Dennoch muss gleichzeitig im Auge behalten werden, dass die Anteile für die grenzwertige Kategorie – vor allem bei Jungen und in der Gesamtstichprobe – bei Anwendung der für diese Schulstichprobe empfohlenen Grenzwerte eher bei 20% liegen, während die kombinierten Anteile für die grenzwertige und auffällige Kategorie von wenigen Ausnahmen abgesehen zum Teil weit über die angestrebten 20% hinausgehen und für die beiden Skalen *Hyperaktivität* und *Prosoziales Verhalten* in beiden Altersgruppen zusammen (Gesamtgruppe) bei bzw. über 30% betragen. Dieses Verteilungsmuster würde allerdings eine Diskrepanz zwischen den von Goodman (1997) empfohlenen Zielvorgaben von 80–10–10% und den zu deren Annäherung vorgeschlagenen neuen Grenzwerten in dieser Studie anzeigen und wäre ein Hinweis darauf, dass die Grenzwerte zu weich gewählt wurden, da dadurch ein zu hoher Anteil der Kinder als grenzwertig klassifiziert wird. Die Grenzwerte wurden in dieser Studie jedoch bewusst trotzdem so gewählt, da die vorliegende Studie eine höhere Sensitivität beim Identifizieren von Problemfällen als Ziel fokussierte. Dies geht allerdings zwangsläufig immer zu Lasten der Spezifität und führt in der Folge dazu, dass nur sehr wenige Kinder und Jugendliche auf allen fünf SDQ-Subskalen mit ihren Rohwerten unterhalb des grenzwertigen Bereichs verbleiben.

Die Entscheidung darüber, ob Sensitivität oder Spezifität zu bevorzugen ist, lässt sich jedoch nicht eindeutig festlegen und muss sich immer nach der speziellen Fragestellung und den institutionellen Rahmenbedingungen richten. Besteht das Ziel einer Studie eher in einer geringeren Anzahl von grenzwertigen Kindern – in der Gefahr, Probleme dabei zu übersehen – sollte eine entsprechende Anpassung der vorgeschlagenen Grenzwerte mit den Ta-

bellens der Rohwertverteilung im elektronischen Supplement 2 dieser Studie vorgenommen werden. Zur Auswertung des SDQ-S wird empfohlen, deshalb nur die Kategorie *auffällig* als wirklich kritisch anzusehen. Das heißt: Ein im SDQ-S als *grenzwertig* eingestuftes Ergebnis sollte zunächst nur als Hinweis auf ein problematisches Verhalten gewertet werden und mit spezifischeren Instrumenten für den jeweiligen Bereich noch einmal überprüft werden. Zu berücksichtigen ist zudem, dass die in dieser Studie vorgeschlagenen Grenzwerte alters- und geschlechtsspezifisch bestimmt wurden, während die britischen Grenzwerte für die SDQ-Selbstbeurteilungsversion von Goodman et al. (1998) – wahrscheinlich aufgrund der geringen Stichprobengröße – alters- und geschlechtsunspezifisch für die Altersgruppe 11;0–17;11 Jahren festgelegt wurden. Dass jedoch eine alters- und geschlechtsspezifische Anpassung von Grenzwerten angemessen ist, unterstreichen in dieser Studie nicht zuletzt die hohen kombinierten Prozentanteile der Kategorien *grenzwertig* und *auffällig* für die beiden SDQ-Skalen *Hyperaktivität* und *Emotionale Probleme*. Der höchste kombinierte Prozentanteil ergab sich bei Jungen sowohl in der jüngeren als auch in der älteren Altersgruppe für die Skala *Hyperaktivität*, wobei der Prozentanteil in der älteren Altersgruppe um ca. 1 % Prozent höher lag als in der jüngeren Altersgruppe.

Auch bei Mädchen ließ sich eine Zunahme für diese SDQ-Skala in der älteren Altersgruppe konstatieren und in der jüngeren Altersgruppe fiel der höchste Prozentanteil auf die SDQ-Skala *Emotionale Probleme*. Der letztgenannte Befund lässt sich auch in die Ergebnisse anderer Studien einreihen, denen zufolge Mädchen häufiger emotionale Probleme berichten als Jungen (Bilz, 2014; Hampel & Petermann, 2006; Schmitz, Vierhaus & Lohaus, 2012). Dazu passen unter anderem die Befunde von Kullik und Petermann (2013) oder von Koglin, Petermann, Jaščenoka, Petermann und Kullik (2013), die auf geschlechtsspezifische Unterschiede in der Emotionsregulation hindeuten. Auch die überwiegend deutlich geringeren kombinierten Prozentanteile in den vier weiteren SDQ-Problemskalen bei Mädchen sind nicht ungewöhnlich, da Mädchen in der Regel weniger Verhaltensprobleme als Jungen zeigen (Hölling et al., 2014; Moffitt, Caspi, Rutter & Silva, 2001; Petermann & Petermann, 2013).

Die vorliegende Studie weist einige Limitationen auf, die in zukünftigen Studien zu beachten sind: Darauf hinzuweisen ist zunächst, dass alle berichteten Ergebnisse auf einer sehr altersheterogenen und regionalen Schulstichprobe basieren, in der die Mehrzahl der Schülerinnen und Schüler aus Sekundarschulen stammen ($n = 1\,141$) und sich nur wenige Gymnasiasten befinden ($n = 360$). Dies bedeutet schließlich eine systematische Verzerrung aller mit der Gesamtstichprobe gewonnenen Ergebnisse, da die Höhe und Verteilung von SDQ-Skalenwerten bekannterweise sowohl vom Alter und Geschlecht als auch von der

sozioökonomischen Schicht der Schülerinnen und Schüler und damit letztlich von dem besuchten Schulzweig abhängen. In klinischen und kinderpsychiatrischen oder auch altershomogeneren Stichproben dürften sich demzufolge andere Grenzwerte ergeben. Zudem muss daran erinnert werden, dass zur Bestimmung der Grenzwerte in dieser Studie lediglich zwischen zwei Altersgruppen differenziert wurde, sodass die vorliegenden Ergebnisse eventuell auch mit allgemeinen Entwicklungseffekten konfundiert sein könnten. Direkte Vergleiche mit Grenzwerten, Skaleneigenschaften etc. aus anderen Studien, die auf Daten altershomogenerer Samples basieren, sollten daher nur mit gewisser Vorsicht vorgenommen werden. Aufgrund des begrenzten Umfangs in einigen Altersgruppen bzw. der regionalen Stichprobe sowie der unterschiedlichen Zusammensetzungen der einzelnen Alters- und Geschlechtsgruppen war eine altersspezifischere Analyse in dieser Studie jedoch nicht möglich. Weitere Studien müssen deshalb altershomogenere Samples untersuchen. Da sich in der vorliegenden Studie zudem auf mehreren SDQ-Skalen signifikante Geschlechts- und Schulformeffekte abbildeten, die sich jedoch aufgrund der regionalen Stichprobe nicht hinreichend stichprobenspezifisch begründen lassen, muss auf eine Neubewertung des SDQ in dieser Studie verzichtet werden.

Zu diskutieren wäre weiterhin generell über den Nutzen von geschlechts- und altersspezifischen SDQ-Grenzwerten, da die von Goodman (1997) festgelegten Zielvorgaben nur ein an Prävalenzraten orientiertes und willkürlich vorab festgelegtes alters- und geschlechtsübergreifendes Grenzwert-Kriterium darstellen. Insgesamt verdeutlichen schließlich auch die Ergebnisse in der vorliegenden Studie, dass die britischen Grenzwerte von Goodman et al. (1998) nicht unmittelbar auf andere Stichproben übertragbar sind, was auch schon Altendorfer-Kling et al. (2007) in ihrer Studie berichten. Zu beachten ist dabei, dass kinderpsychiatrische und klinische Prävalenzraten in der Regel (sowohl insgesamt als auch diagnosespezifisch) systematisch je nach Alter und Geschlecht variieren und die Skalenmittelwerte je Alter und Geschlecht idealerweise die epidemiologisch quantifizierbaren klinischen Prävalenzraten widerspiegeln. Daraus lässt sich allerdings der Schluss ziehen, dass eine geschlechtsspezifische Beurteilung der in einer Normierungsstichprobe erhaltenen SDQ-Skalenrohwerte eventuell weniger funktional ist, wenn verschiedene Grenzwerte für Jungen und Mädchen berechnet werden, wonach jeweils 10% aller Mädchen und Jungen als auffällig klassifiziert werden. So werden durch ein solches Vorgehen klinisch auffällige Jungen nur zum Teil als auffällig diagnostiziert, während Mädchen häufiger zu Unrecht als auffällig klassifiziert werden. Ein weiterer Nachteil der SDQ-Grenzwerte ist zudem, dass sich die a priori vorgegebenen Prozentanteile von 80–10–10% besonders bei den fünf SDQ-Subskalen aufgrund der geringen Itemanzahl und den damit einhergehenden wenig möglichen

Rohwertstufen (0–10) weniger genau erzielen lassen. Bei einer Annäherung an die im SDQ festgelegten prozentualen Zielvorgaben von 80–10–10% treten somit in den vorliegenden, getrennt normierten Untergruppen Ungenauigkeiten auf, da man sich für empirisch beobachtete Anteile entscheiden muss, die entweder ein wenig über oder ein wenig unter diesen Zielvorgaben liegen. SDQ-Grenzwerte dürfen daher immer nur allgemein als grobe Richtwerte betrachtet werden. Um das Vorliegen einer klinisch relevanten Störung vorherzusagen, könnten eventuell genauso gut die individuellen Rohwerte der fünf SDQ-Subskalen je nach Geschlecht oder Alter herangezogen werden (vgl. Goodman & Goodman, 2011). Damit ergibt sich für die vorliegende Studie letztlich die Schlussfolgerung, dass die hier empfohlenen Grenzwerte die bekannten alters- und geschlechtsabhängigen Differenzen in klinischen Prävalenzraten nivellieren. Zur Quantifizierung und Dokumentation von Alters- oder Geschlechtsunterschieden in mittleren SDQ-Subskalenroherten sollten deshalb immer auch die Skalenmittelwerte in Abhängigkeit von der jeweiligen untersuchten Altersgruppe und dem Geschlecht berücksichtigt werden, wie dies in der vorliegenden Studie zusätzlich mit einer grafisch beschreibenden Darstellung von SDQ-auffälligen und -grenzwertigen Prozentanteilen je Altersgruppe und Geschlecht veranschaulicht wurde. Um ein klinisch auffälliges Verhalten bei Kindern und Jugendlichen zu diagnostizieren, sollten vor allem auch möglichst verschiedene Informationsquellen unter Zuhilfenahme von Eltern- oder Lehrkräfteeinschätzungen herangezogen und miteinander kontrastiert werden, um z. B. die nicht selten zu beobachteten Verzerrungseffekte von Selbst- und Fremdbeurteilungen erkennen zu können.

Ungeklärt bleibt in dieser Studie zudem die Frage nach der Validität des SDQ-Deu-S. Da eine hinreichende Validität jedoch erst die Anwendbarkeit eines Verfahrens rechtfertigt, müssen weitere Studien – wie beispielsweise die von Becker et al. (2004) – die Validität des SDQ-Deu-S umfassend prüfen. Wenngleich die eher geringe Reliabilität der fünf SDQ-Subskalen als problematisch anzusehen wäre, soll abschließend noch einmal betont werden, dass bei der Evaluation von einzelnen SDQ-Versionen auch nicht nur ein einziges Instrument vorliegt, sondern drei verschiedene Versionen – für Eltern, für Lehrkräfte bzw. Erziehende und die der Selbsturteile von Kindern und Jugendlichen – existieren, in denen die gleichen Iteminhalte erfragt werden und zu einer einheitlichen Skalenbildung herangezogen werden können. Auch wenn die bisher weltweit damit angestellten Evaluationen psychometrischer Eigenschaften (inklusive der faktoriellen Struktur) nicht für alle drei Versionen in jedem untersuchten Altersbereich und jeder Stichprobenszusammensetzung zu exakt den gleichen positiven Ergebnissen kommen, sind punktuelle Veränderungen (z. B. für nur eine der drei Versionen oder für einen bestimmten engen Altersbereich, ein Geschlecht, eine nationale Übersetzung) an der insgesamt gut replizierbaren und ausreichend

validierten Skalenstruktur des Instruments nicht zu empfehlen. Zusammenfassend lässt sich vor diesem Hintergrund damit festhalten, dass der SDQ-Deu-S im Vergleich zur CBCL ein deutlich ökonomischeres und gut anwendbares Screening-Verfahren für Kinder und Jugendliche ist, das die Verhaltensprobleme und die Verhaltensstärken von Kindern und Jugendlichen global erfassen kann.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar über <http://dx.doi.org/10.1026/0012-1924/a000153>

ESM 1. Grafische Darstellungen der Skalenmittelwerte je Geschlecht und Altersgruppe, der prozentualen Verteilung von auffälligen und grenzwertigen Jugendlichen in dieser Stichprobe bei Auswertung nach britischen Grenzwerten sowie der prozentualen Verteilung von auffälligen und grenzwertigen Jugendlichen nach in dieser Studie vorge schlagenen Grenzwerten.

ESM 2. Rohwerteverteilung in der deutschen Selbstbeurteilungsversion des Strengths and Difficulties Questionnaire (SDQ-Deu-S) bei Jungen ($n = 845$) und Mädchen ($n = 794$).

Literatur

- Altendorfer-Kling, U., Ardel-Gattinger, E. & Thun-Hohenstein, L. (2007). Der Selbstbeurteilungsbogen des SDQ anhand einer österreichischen Feldstichprobe. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 35, 265–271.
- Arbeitsgruppe Deutsche Child Behavior Checklist (2012a). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen – Revision 2012 (CBCL/4–18 R)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).
- Arbeitsgruppe Deutsche Child Behavior Checklist (2012b). *Lehrerfragebogen über das Verhalten von Kindern und Jugendlichen – Revision 2012 (TRF-R)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).
- Arbuckle, J. L. (2012). *AmosTM 21.0 User's Guide*. Chicago: SPSS.
- Becker, A., Hagenberg, N., Roessner, V., Woerner, W. & Rothenberger, A. (2004). Evaluation of the self-reported SDQ in a clinical setting: do self-reports tell us more than ratings by adult informants? *European Child and Adolescent Psychiatry*, 13, 17–24.
- Becker, A., Steinhausen, H. C., Balursson, G., Dalsgaard, S., Lorenzo, M. J., Ralston, S. J. et al. (2006). Psychopathological screening of children with ADHD: Strengths and Difficulties Questionnaire in a pan-European study. *European Child and Adolescent Psychiatry*, 15, 56–62.
- Bilz, L. (2014). Werden Ängste und depressive Symptome von Kindern und Jugendlichen in der Schule übersehen? *Zeitschrift für Pädagogische Psychologie*, 28, 57–62.
- Borg, A. M., Kaukonen, P., Salmelin, R., Joukamaa, M. & Tamminen, T. (2012). Reliability of the Strengths and Dif-

- facilities Questionnaire among Finnish 4–9-year-old children. *Nordic Journal of Psychiatry*, 66, 403–413.
- Capron, C., Théron, C. & Duyme, M. (2007). Psychometric properties of the French version of the self-report and teacher Strengths and Difficulties Questionnaire (SDQ). *European Journal of Psychological Assessment*, 23, 79–88.
- Dicky, W. C. & Blumberg, S. J. (2004). Revisiting the factor structure of the Strengths and Difficulties Questionnaire: United States, 2001. *Journal of American Academy of Child and Adolescent Psychiatry*, 43, 1159–1167.
- Essau, C. A., Olaya, B., Anastassiou-Hadjicharalambous, X., Pauli, G., Gilvarry, C., Bray, D. et al. (2012). Psychometric properties of the Strengths and Difficulties Questionnaire from five European countries. *International Journal of Methods Psychiatric Research*, 21, 232–245.
- Giannakopoulos, G., Dimitrakaki, C., Papadopoulou, K., Tzavara, C., Kolaitis, G., Ravens-Sieberer, U. et al. (2013). Reliability and validity of the Strengths and Difficulties Questionnaire in Greek adolescents and their parents. *Health*, 5, 1774–1783.
- Gómez-Beneyto, M., Nolasco, A., Moncho, J., Pereyra-Zamora, P., Tamayo-Fonseca, N., Munarriz, M. et al. (2013). Psychometric behaviour of the Strengths and Difficulties Questionnaire (SDQ) in the Spanish national health survey 2006. *BMC Psychiatry*, 13, Art 95.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire. A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, 40, 1337–1345.
- Goodman, R. (2005). Strengths and Difficulties Questionnaire. Information for researchers and professionals about the Strengths & Difficulties Questionnaires. Retrieved May 6, 2015, from <http://www.sdqinfo.org/py/sdqinfo/b3.py?language=German>
- Goodman, A. & Goodman, R. (2011). Population mean scores predict child mental disorder rates: validating SDQ prevalence estimators in Britain. *Journal of Child Psychology and Psychiatry*, 52, 100–108.
- 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, 1179–1191.
- Goodman, R., Meltzer, H. & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 7, 125–130.
- Hampel, P. & Petermann, F. (2006). Perceived stress, coping, and adjustment in adolescents. *Journal of Adolescent Health*, 38, 409–415.
- He, J. P., Burstein, M., Schmitz, A. & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): the factor structure and scale validation in U.S. adolescents. *Journal of Abnormal Child Psychology*, 41, 583–595.
- Hölling, H., Schlack, R., Petermann, F., Ravens-Sieberer, U., Mauz, E. & KiGGS Study Group (2014). Psychische Auffälligkeiten und psychosoziale Beeinträchtigungen bei Kindern und Jugendlichen im Alter von 3 bis 17 Jahren in Deutschland. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 57, 807–819.
- Kóbor, A., Takás, A. & Urbán, R. (2013). The bifactor model of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 29, 299–307.
- Koglin, U., Barquero, B., Mayer, H., Scheithauer, H. & Petermann, F. (2007). Deutsche Version des Strengths and Difficulties Questionnaire (SDQ-Deu): Psychometrische Qualität der Lehrer-/Erzieherversion für Kindergartenkinder. *Diagnostica*, 53, 175–183.
- Koglin, U., Petermann, F., Jaščenoka, J., Petermann, U. & Kullik, A. (2013). Emotionsregulation und aggressives Verhalten im Jugendalter. *Kindheit und Entwicklung*, 22, 155–164.
- Koskelainen, M., Sourander, A. & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *European Child and Adolescent Psychiatry*, 10, 180–185.
- Kullik, A. & Petermann, F. (2013). Attachment to parents and peers as a risk factor for adolescent depressive disorders: the mediating role of emotion regulation. *Child Psychiatry and Human Development*, 44, 537–548.
- Laakmann, M., Schultheiß, J., Petermann, F. & Petermann, U. (2013). Zur Wirksamkeit des JobFit-Trainings – ein Vergleich mit und ohne Migrationshintergrund. *Zeitschrift für Psychiatrie, Psychologie und Psychotherapie*, 61, 189–196.
- Lai, K. Y., Luk, E. S., Leung, P. W., Wong, A. S., Law, L. & Ho, K. (2010). Validation of the Chinese version of the Strengths and Difficulties Questionnaire in Hong Kong. *Social Psychiatry and Psychiatric Epidemiology*, 45, 1179–1186.
- Liu, S. K., Chien, Y. L., Shang, C. Y., Lin, C. H., Liu, Y. C. & Gau, S. S. F. (2013). Psychometric properties of the Chinese version of Strengths and Difficulties Questionnaire. *Comprehensive Psychiatry*, 54, 720–730.
- Lohbeck, A., Nitkowski, D., Petermann, F. & Petermann, U. (2014). Erfassung von Schülerelbstschätzungen zum schulbezogenen Sozial- und Lernverhalten: Validierung einer Schülereinschätzliste für Sozial- und Lernverhalten (SSL). *Zeitschrift für Erziehungswissenschaft*, 17, 701–722.
- Mansbach-Kleinfeld, I., Apter, A., Farbstein, I., Levine, S. Z. & Ponizovsky, A. M. (2010). A population-based psychometric validation study of the Strengths and Difficulties Questionnaire – hebrew version. *Frontiers in Psychiatry*, 1, 151.
- Mieloo, C. L., Bevaart, F., Donker, M. C., van Oort, F. V., Raat, H. & Jansen, W. (2014). Validation of the SDQ in a multi-ethnic population of young children. *European Journal of Public Health*, 24, 26–32.
- Moffitt, T. E., Caspi, A., Rutter, M. & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge, NY: Cambridge University Press.
- 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, 437–448.
- Palmieri, P. A. & Smith, G. C. (2007). Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a U.S. sample of custodial grandmothers. *Psychological Assessment*, 19, 189–198.
- Petermann, F. & Petermann, U. (2013). Störungen des Sozialverhaltens. *Kindheit und Entwicklung*, 22, 123–126.
- Richter, J., Sagatun, A., Heyerdahl, S., Oppedal, B. & Røysamb, E. (2011). The Strengths and Difficulties Questionnaire (SDQ) – self-report. An analysis of its structure in a

- multiethnic urban adolescent sample. *Journal of Child Psychology and Psychiatry*, 52, 1002–1011.
- Riso, D. D., Salcuni, S., Chessa, D., Raudino, A., Lis, A. & Altoè, G. (2010). The Strengths and Difficulties Questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Differences*, 49, 570–575.
- Ruchkin, V., Kopusov, R., Vermeiren, R. & Schwab-Stone, M. (2012). The Strengths and Difficulties Questionnaire: Russian validation of the teacher version and comparison of teacher and student reports. *Journal of Adolescence*, 35, 87–96.
- Schmitz, A. K., Vierhaus, M. & Lohaus, A. (2012). Geschlechtstypische Unterschiede und geschlechtstypische Erwartungen beim Einsatz von Bewältigungsstrategien und ihre Zusammenhänge zum Problemverhalten von Jugendlichen. *Zeitschrift für Gesundheitspsychologie*, 20, 13–21.
- Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A. & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4-to 12-year-olds: A review. *Clinical Child and Family Psychology Review*, 13, 254–274.
- Strengths and Difficulties Questionnaire (2015a). Information for researchers and professionals about the Strengths & Difficulties Questionnaires. Retrieved May 6, 2015, from <http://www.sdqinfo.com/>
- Strengths and Difficulties Questionnaire (2015b). Information for researchers and professionals about the Strengths & Difficulties Questionnaires. Retrieved May 6, 2015, from <http://www.sdqinfo.org/py/sdqinfo/b3.py?language=German>
- 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.
- Weiber, R. & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS* (2., erweit. Aufl.). Heidelberg: Springer.
- Woerner, W., Becker, A., Friedrich, C., Klasen, H., Goodman, R. & Rothenberger, A. (2002). Normierung und Evaluation der deutschen Elternversion des Strengths and Difficulties Questionnaire (SDQ): Ergebnisse einer repräsentativen Felderhebung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 30, 105–112.
- Yao, S., Zhang, C., Zhu, X., Jing, X., McWhinnie, C. M. & Abela, J. R. (2009). Measuring adolescent psychopathology: psychometric properties of the self-report Strengths and Difficulties Questionnaire in a sample of Chinese adolescents. *Journal of Adolescent Health*, 45, 55–62.

Dr. Annette Lohbeck

Carl von Ossietzky Universität Oldenburg
Institut für Pädagogik
Uhlhornsweg 82
26129 Oldenburg
E-Mail: annette.lohbeck@uni-oldenburg.de

Dipl.-Psych. Jan Schultheiß
Prof. Dr. Franz Petermann
Prof. Dr. Ulrike Petermann

Universität Bremen
Zentrum für Klinische Psychologie
und Rehabilitation (ZKPR)
Grazer Straße 6
28359 Bremen