

Empirische Sonderpädagogik, 2019, Nr. 2, S. 160-171
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Testgüteuntersuchungen des Fragebogens „Soziale Distanz in inklusiven Settings“ (SoDiS). Validität und Test-Retest-Reliabilität

Anita Gerullis & Christian Huber

Bergische Universität Wuppertal

Zusammenfassung

Diese Studie untersucht die Itemcharakteristika, die Validität und Reliabilität des Fragebogenverfahrens zur Messung Sozialer Distanz in inklusiven Settings (SoDiS, Gerullis & Huber, 2016). Das Verfahren beinhaltet vier bildliche Items, die dem Alltag der Kinder nachempfunden sind (Klassenraum, Schulhof, Essensraum, Kinosaal) und wenig sprachliche Kompetenzen verlangen. Kinder mit Auffälligkeiten (geistiger Behinderung, Lernbehinderung, Auffälligkeit im Verhalten (ADHS), körperlicher Behinderung und Migrationshintergrund) werden dabei anhand von Bild- und Text-Vignetten dargestellt und sollen von Probandinnen und Probanden in Bezug auf die Nähe der Interaktion geratet werden. In einem klassischen Paper-Pencil-Fragebogendesign bearbeiteten $N = 617$ Schülerinnen und Schüler zu zwei Messzeitpunkten im Abstand von vier Wochen den SoDiS. Zur Prüfung der konvergenten Validität wurde die CATCH-Skala (Schwab, 2015a) herangezogen. Es zeigten sich mittlere bis hohe Itemtrennschärfen ($r_{it} = .481$ bis $r_{it} = .684$) und eine gute konvergente Validität (Korrelation der Gesamtskalen: $r = .73$, $p < .01$). Die Gesamtskala des SoDiS weist eine gute Test-Retest-Reliabilität von $r = .83$ ($p = .01$) auf. Die Testgüteuntersuchungen zeigen somit insgesamt gute Ergebnisse.

Schlüsselwörter: Soziale Distanz, Messung, Inklusion, Validität, Test-Retest-Reliabilität

Examining the test criteria of the questionnaire “Social Distance in Inclusive Settings” (SoDiS). Validity and test-retest-reliability.

Abstract

This study examines the item characteristics, the validity and the reliability of the SoDiS questionnaire measuring “social distance in inclusive settings” (Gerullis & Huber, 2016). This questionnaire consists of four pictorial items which portray situations of relevance to school children (classroom, school compound, cafeteria, and cinema) and requires only little reading competences. Children with diverse disabilities (intellectual disability, learning disability, behavioral disorder (ADHD), and physical impairment) and a child with a migrant background are shown in pictorial and textural vignettes and subjects rate their desired closeness to these children in the given situations. $N = 617$ students were surveyed twice within a four week time period using the SoDiS and the CATCH-Scale (Schwab, 2015a). We were able to show medium to high discriminatory power of the four items ($r_{it} = .481$ to $r_{it} = .684$) and good convergent validity (correlations of the total scales: $r = .73$, $p < .01$). Our data indicated good test-retest-reliability

($r = .83$, $p = .01$). Overall the SoDiS shows good test criteria.

Keywords: social distance, measure, inclusive education, validity test-retest-reliability

Einleitung

Inklusive Strukturen und soziale Distanz

Kinder mit sonderpädagogischem Förderbedarf zeigen in inklusiven Settings zwar bessere schulische Leistungen (Feyerer, 1998; Lindsay, 2007; Myklebust, 2002; Preuss-Lausitz, 2009; Wocken, 1999), sind in ihrer Klassengemeinschaft jedoch oft nicht genügend integriert oder werden von den Mitschülerinnen und Mitschülern abgelehnt (Bakker & Bosman, 2003; Bless & Mohr, 2007; Bossaert et al., 2012; Frederickson & Funham, 2001; Haerberlin, 1991; Huber, 2008; Huber & Wilbert, 2012; Kavale & Forness, 1996; Lindsay, 2007; Ruijs, Peetsma & Van der Veen, 2010; Schwab, 2015a; Schwab, 2015b). Ein geeignetes Konstrukt zur Untersuchung sozialer Ausgrenzung in der Schule ist die *soziale Distanz* (Gerullis & Huber, 2016; 2017). Sie beschreibt die Ablehnung einer Person aufgrund der Zuschreibung einer vermuteten Gruppenzugehörigkeit und besteht aus einer interaktiven, einer normativen und einer affektiven Komponente (Gerullis & Huber, 2016; 2017). Im Gegensatz zu klassischen soziometrischen Verfahren (Buckowski, Cillessen & Velásquez, 2012; Cillessen, 2011; Huber, 2006; Koster, Nakken, Pijl & van Houten 2009; Moreno, 1974) wird nicht das Verhältnis zu einer bestimmten Mitschülerin / einem bestimmten Mitschüler betrachtet, sondern die Bereitschaft mit einem Mitglied einer sozialen Gruppe in Kontakt zu treten beschrieben – unter Einbezug eines bewertenden Aspekts und einem subjektiven Gefühl, welches in der Interaktion mitschwingt (Gerullis & Huber, 2017). Bisherige Untersuchungen zeigen, dass die soziale Distanz ein kontextabhängiges Phäno-

men ist und je nach Situation (soziale Situation vs. Leistungssituation) unterschiedlich berichtet wird (Gasser et al., 2012; Laws & Kelly, 2005; Nowicki, 2006). Bisherige Verfahren zur Messung sozialer Distanz (Aboud & Mitchell, 1977; Bogardus, 1933; Bossaert et al., 2013; Catapusan, 1954; Clack, 2007; Connolly, 2006; Greenwald, 1998; Holzkamp, 1964; Lauber, 2004; Rosenbaum, Armstrong & King, 1985; Schwab, 2015a; Verna, 1981; Wocken, 1992) stammen zu großen Anteilen aus der Rassismusforschung und weisen einen mangelnden Bezug zur schulischen Inklusion auf. Vorhandene Testverfahren weisen nicht nur testtheoretische Mängel auf, sondern sind zudem aufgrund anspruchsvoller sprachlicher Formulierungen für den Einsatz in der Primarstufe nicht geeignet (Gerullis & Huber, 2016; 2017).

Die Messung sozialer Distanz mit dem SoDiS und bisherige Validierung des Verfahrens

Der SoDiS (Soziale Distanz in inklusiven Settings) (Gerullis & Huber, 2016) ist ein Fragebogenverfahren zur Untersuchung der sozialen Distanz von Grundschulkindern zu Kindern mit Behinderungen und Migrationshintergrund (*eine genauere Beschreibung des Testverfahrens befindet sich im Methodenteil*). Er ist aufgrund des Einsatzes von leicht verständlichen Vignetten (Kinder mit Lernbehinderung, geistiger Behinderung, körperlicher Behinderung, Verhaltensauffälligkeit, Migrationshintergrund) und alltagsnahen und anschaulichen Items gut für den Einsatz im schulischen Primarbereich geeignet. Gerullis und Huber (2016) konnten zeigen, dass der SoDiS genügend Antwortvarianz aufweist und dass die Items keine Boden- oder Deckeneffekte zeigen. Die Items weisen eine mittlere Inter-

Item-Korrelation und eine gute Trennschärfe auf. Im Rahmen der Validitätsprüfung zeigten sich gute Werte in der konvergenten Validität und eine eindimensionale faktorielle Struktur. Im Hinblick auf die Konstruktvalidierung berichten Gerullis und Huber (2016) von mit bisherigen Forschungsergebnissen übereinstimmenden Zusammenhängen der sozialen Distanz mit dem Alter, dem Geschlecht und der sozialen Situation (De Boer et al., 2012; Gasser et al., 2012; Schwab, 2015a). Der SoDiS erweist sich als ein valides und praktikables Fragebogenverfahren für den schulischen inklusiven Kontext. Amelang und Schmidt-Atzert (2006) und Bühner (2011) beschreiben jedoch vor allem die Reliabilität als ein zentrales Gütekriterium für psychologische Messungen und stellen hier die Test-Retest-Methode bei von Übungseinflüssen unabhängigen Testverfahren in den Vordergrund. Der SoDiS wurde bisher nicht auf seine Test-Retest-Reliabilität untersucht.

Anknüpfung an die vorangehende Studie und Anliegen dieser Studie

Die hier beschriebene Studie zielt neben einer Untersuchung der Itemcharakteristika und der Validität auch auf eine Reliabilitätsprüfung des SoDiS ab und knüpft damit an eine vorangehende Studie zur Testgüte des SoDiS (Gerullis & Huber, 2016) an: In dieser vorangehenden Studie zeigten sich eine akzeptable bis gute Skalenhomogenität ($\alpha = .68$) und akzeptable bis gute Trennschärfen ($r_{it} = .25$ bis $r_{it} = .49$). Die konvergente Validität konnte anhand mittlerer Korrelationen mit zwei Außenkriterien (Childrens' Social Distance Scale, Connolly, 2006 ($\alpha = .66, p = 0.1$) und der Kurzversion der CATCH-Skala, Schwab, 2015 ($\alpha = .72, p = 0.1$)) gezeigt werden.

Das erste Ziel dieser Studie ist die Sicherung der Ergebnisse der vorherigen Studie zu den Itemcharakteristika und der Validität des SoDiS-Verfahrens (Gerullis & Huber, 2016). Lüke & Grosche (2018) verdeutlichen die Notwendigkeit der Replikation

von Ergebnissen zur Testgüteuntersuchungen im Feld der Einstellungsforschung im inklusiven Gegenstandsbereich. Sie beschreiben, dass im deutschen Sprachraum kaum Publikationen zu finden sind, die sich einer erneuten Prüfung der Ergebnisse unterzogen und theoretische Modelle und Kennwerte an neuen Stichproben untersucht haben. Zudem wurden in dieser zweiten Studie erstmalig auch Zweitklässlerinnen und Zweitklässler in die Untersuchung mit einbezogen. Das zweite Ziel der Studie ist die Prüfung der Test-Retest-Reliabilität der Items durch einen Einsatz des Verfahrens zu zwei Messzeitpunkten. Hieraus ergeben sich folgende Fragestellungen:

1. Wie ist die Ausprägung der sozialen Distanzen zu den einzelnen Fallvignetten und welche Unterschiede stellen sich zwischen den Messzeitpunkten dar?
2. Können bisherige Ergebnisse zu den Itemcharakteristika repliziert werden?
3. Können bisherige Ergebnisse zur konvergenten Validität repliziert werden?
4. Weisen die Items des SoDiS eine ausreichende Test-Retest-Reliabilität auf?

Im Hinblick auf die Ausprägung der sozialen Distanzen erwarten wir signifikant höhere Distanzen zu Kindern mit Auffälligkeiten als zu Kindern ohne Auffälligkeiten. Im Sinne der Replikation der Untersuchung der Itemcharakteristika und der konvergenten Validität erwarten wir mit der vorherigen Studie vergleichbare Ergebnisse (s.o.). Wir erwarten, dass die Items des SoDiS gut zwischen Personen mit hoher und niedriger Merkmalsausprägung unterscheiden können (Trennschärfen $r_{it} > .3$) und eine homogene Skala abbilden ($\alpha > .8$). In Bezug auf die konvergente Validität erwarten wir eine signifikante Korrelationen mit dem Außenkriterium (CATCH-Skala, Schwab, 2015) von $r > .6$. Im Bereich der Test-Retest-Reliabilität wird eine mittlere Korrelation der Werte zwischen den Messzeitpunkten ($r > .8$) (Wertebereiche in Anlehnung an Bühner, 2011) erwartet.

Methoden

Stichprobe und Design

Zielgruppe dieser Untersuchung waren Schülerinnen und Schüler in Nordrhein-Westfalen im 2. bis 6. Schulbesuchsjahr. Zufällig ausgewählte Schulen wurden telefonisch kontaktiert und erhielten bei Interesse an der Studienteilnahme weitere Informationen per E-Mail. Einbezogen wurden nur Regelschulen, an denen auch Kinder mit Förderbedarfen beschult werden. Das Studiendesign entspricht dem Design einer einfachen Fragebogenstudie im Feld mit zwei Messzeitpunkten (MZP). Insgesamt nahmen $N = 634$ Schülerinnen und Schüler aus den Klassenstufen zwei bis sechs aus 13 Schulen an der Befragung teil. 50,9% waren weiblich, 49,1% waren männlich. Nur $N = 617$ Teilnehmerinnen und Teilnehmer konnten zu beiden MZP Angaben machen und wurden in der Datenanalyse berücksichtigt. 83 Kindern wurde im Vorfeld der Studie ein sonderpädagogischer Förderbedarf diagnostiziert (Förderbedarf im Bereich Lernen: 49; Sprache: 4; emotional-soziale Entwicklung: 30; Sehen: 1). Das durchschnittliche Alter der Teilnehmerinnen und Teilnehmer betrug $M = 9,1$ Jahre ($SD = 1.23$).

Material

Der SoDiS – Testbogen zur Messung Sozialer Distanz zu Kindern mit Auffälligkeiten

Der SoDiS umfasst vier alltagsnahe und anschauliche Items zur Messung sozialer Distanz von Grundschülerinnen und Grundschulern gegenüber Kindern mit Behinderungen (Kind mit geistiger Behinderung, Kind mit Lernbehinderung, Kind mit körperlicher Behinderung, Kind mit Verhaltensauffälligkeit) und Migrationshintergrund. Diese werden jeweils in einer Bild- und einer Textvignette dargestellt. Zusätzlich wird ein Kind ohne Auffälligkeiten präsentiert. Die-

ses beinhaltet keine Textvignette (Gerullis & Huber, 2016). Zur Bearbeitung des SoDiS stellen Probandinnen und Probanden sich vor, das in der Vignette dargebotene Kind sei neu in der Klasse und markieren, an welcher von fünf freien Stellen das Kind spielen oder sitzen soll. Hierbei werden den Teilnehmerinnen und Teilnehmern grafische Items verschiedener kindgerechter Situationen (Klassenraum, Schulhof, Essensraum, Kinosaal) präsentiert. Zur Auswertung des SoDiS wird pro Vignette ein Wert anhand der Entfernung zum eigenen (Sitz-)platz festgemacht. Die Werte werden von eins (dichtester Platz) bis fünf (weit entferntester Platz) kodiert. Aus der Summe der einzelnen Itemwerte lässt sich pro Vignette ein Gesamtscore errechnen (Gerullis & Huber, 2016).

CATCH-Skala zur Überprüfung der konvergenten Validität

Zur Validierung des SoDiS wurde eine Kurzversion der CATCH-Skala von Rosenbaum, Armstrong und King (1985) verwendet (Schwab, 2015a). Diese Skala operationalisiert die soziale Distanz anhand von Items zur Kontaktbereitschaft mit dem jeweiligen Kind. Sie besteht aus sechs Items und wurde auf ihre Testgüte geprüft (Bossaert & Petry, 2013; Rosenbaum et al., 1985; Schwab, 2015a). Die CATCH-Skala beinhaltet ein fünfstufiges Antwortformat. Grundlage für das Rating bildeten die Vignetten aus dem SoDis.

Soziodemografische Variablen

Im Schülerinnen und Schüler-Fragebogen machten die Teilnehmerinnen und Teilnehmer Angaben zu ihrem Geschlecht, ihrem Alter, ihrer Herkunft und ihren Vorerfahrungen mit Kindern mit Behinderungen. Über einen Lehrkraftfragebogen erfassten wir vor der Durchführung der Klassentestungen Angaben dazu, ob die Kinder einen sonderpädagogischen Förderbedarf aufweisen und um welchen Förderbedarf es sich handelt.

Untersuchungsablauf

Die Erhebungen fanden im Paper-Pencil-Format im Klassenverband zu zwei Messzeitpunkten mit einem Abstand von vier Wochen statt. Die Testleiterinnen und Testleiter wurden vor Untersuchungsbeginn in der Anwendung der Fragebögen geschult und setzten standardisierte Manuale zur Bearbeitung der Fragebögen ein. Um lese-schwachen Schülerinnen und Schülern eine Teilnahme zu ermöglichen, wurden die Items nacheinander von der Testleitung vorgelesen. Die Befragung dauerte je nach Klassenstufe zwischen 45 und 60 Minuten, wobei die Kinder mit Förderbedarf das Tempo vorgaben. Die Fragebögen lagen entsprechend in Jungen- und Mädchenversion vor. In allen Fragebögen wurden die Namen der Schülerinnen und Schüler mit Codenummern versehen und die Kinder wurden auf die Anonymität der Erhebung hingewiesen. Zu Beginn der Erhebung wurden die Teilnehmerinnen und Teilnehmer ausdrücklich darauf hingewiesen, dass es keine falschen oder richtigen Antworten gibt und sie nach ihrem Gefühl antworten sollten. In Absprache mit den Klassenlehrkräften wurde darauf geachtet, dass zwischen dem ersten und zweiten MZP keine Intervention oder Thematisierung von relevanten Inhalten im fortlaufenden Unterricht stattfand, die das Antwortverhalten der Schülerinnen und Schüler beeinflussen könnte.

Daten und Analyseverfahren

Das Hauptaugenmerk dieser Studie lag auf der Replikation der Ergebnisse zu Itemcharakteristika und zur konvergenten Validität aus einer vorangehenden Studie (Gerullis & Huber, 2016) und der Prüfung des SoDiS in Bezug auf die Test-Retest-Reliabilität.

Ausprägung der Werte

Zur Betrachtung der sozialen Distanz Werte wurden Mittelwerte und Standardabweichungen für die einzelnen Fallvignetten pro

Messzeitpunkt berechnet. Im Rahmen eines t-Tests wurden Kinder mit und ohne Auffälligkeiten kontrastiert. Im Rahmen einer Varianzanalyse konnten Haupt- und Interaktionseffekte zwischen einzelnen Fallvignetten und Messzeitpunkten untersucht werden.

Konvergente Validität und interne Konsistenz

Zur Bestimmung der konvergenten Validität wurde die Übereinstimmung des SoDiS-Gesamtscores mit dem Gesamtscore der CATCH-Skala (Schwab, 2015a) zum ersten MZP berechnet. Damit Ergebnisunterschiede zwischen dem SoDiS und der CATCH nicht auf unterschiedliche Vignetten zurückführbar sind, wurde für beide Verfahren jeweils die gleiche Bild- und Text-Vignette verwendet. Laut Lindemann (2016) bezieht sich die CATCH nur „ganz allgemein (...) auf Kinder mit einer Beeinträchtigung“. Daher wurde in der Auswertung im Rahmen der konvergenten Validität die Vignette des Kinds ohne Auffälligkeiten im SoDiS außer Acht gelassen. Die Werte beziehen sich jeweils auf die Gesamtskala des SoDiS (Summe der vier Items). Die interne Konsistenz wurde über einen Cronbach's α -Wert berechnet.

Test-Retest-Reliabilität

Die Test-Retest-Reliabilität errechnete sich entsprechend gängiger Verfahren für Tests zur Erfassung stabiler Merkmale durch die Korrelation der Testwerte aus den beiden Messzeitpunkten (Bühner, 2011; Amelang und Schmidt-Atzert, 2006). Aufgrund der Eindimensionalität des Konstruktes (Gerullis & Huber, 2016) konnten wir Summenwerte der sozialen Distanz für die Auffälligkeiten in den einzelnen Situationen bilden und so die Test-Retest-Reliabilität der Gesamtskala abbilden.

Ergebnisse

Fragestellung 1:

Wie ist die Ausprägung der sozialen Distanzen zu den einzelnen Fallvignetten und welche Unterschiede stellen sich zwischen den Messzeitpunkten dar?

Ausprägung der Werte. Tabelle 1 stellt die durchschnittliche Ausprägung der Ratings der sozialen Distanz zu den einzelnen Fallvignetten dar. In allen Kategorien wurde der Wertebereich von eins (niedrigere soziale Distanz) bis fünf (hohe soziale Distanz) ausgeschöpft. Es zeigen sich keine Boden- oder Deckeneffekte. Die geringste soziale Distanz zeigen die Teilnehmerinnen und Teilnehmer zu beiden MZP zum Kind ohne Auffälligkeiten. Die höchste soziale Distanz besteht im Rating zum Kind mit Verhaltensauffälligkeit (ADHS). In der Durchführung eines t-Tests zur Überprüfung der Signifikanz dieser Mittelwertsunterschiede zeigt sich, dass die durchschnittliche summierte soziale Distanz zu den Kindern der einzelnen Fallvignetten (Kind mit geistiger Behinderung, Kind mit Lernbehinderung, Kind mit körperlicher Behinderung, Kind mit Verhaltensauffälligkeit, Kind mit Migrationshintergrund) signifikant höher ist als zum Kind ohne Auffälligkeiten ($t(614) = 39.52$, $p < .01$). Zur genaueren Untersuchung die-

ser Unterschiede zwischen den einzelnen Fallvignetten haben wir eine Varianzanalyse berechnet. Hier fanden sich nach Bonferroni-Korrektur signifikante Unterschiede zwischen Kindern ohne Auffälligkeiten und der Vignette der Kinder mit Verhaltensauffälligkeit ($F(1,562) = 5,92$, $p < .05$) und geistiger Behinderung ($F(1,562) = 5,71$, $p < .05$), sowie zwischen dem Kind mit Lernbehinderung und dem Kind mit Verhaltensauffälligkeit ($F(1,562) = 5,22$, $p < .05$).

Unterschiede zwischen den Messzeitpunkten. Die Gesamtmittelwerte der sozialen Distanz zu den zwei Messzeitpunkten unterscheiden sich nicht signifikant ($t(562) = -2,47$, $p = .14$). So zeigte sich auch in der Varianzanalyse mit Messwiederholung kein signifikanter Haupteffekt für den Messzeitpunkt ($F(1,562) = 3,26$, $p = .07$). Wie im vorherigen Abschnitt dargestellt, gab es einen Haupteffekt für die Fallvignette ($F(1,562) = 51,77$, $p < .01$). Außerdem zeigte sich ein Interaktionseffekt zwischen dem Messzeitpunkt und der sozialen Distanz zu den einzelnen Fallvignetten ($F(1,562) = 56,38$, $p < .01$).

Tabelle 1: Durchschnittliche Ausprägung der sozialen Distanz zu den zwei Messzeitpunkten (MZP)

Auffälligkeit / Vignette	MZP 1		MZP 2	
	M	SD	M	SD
Kind ohne Auffälligkeit	1.71	1.13	1.75	1.07
Kind mit geistiger Behinderung	2.52	1.36	2.54	1.27
Kind mit Lernbehinderung	1.98	1.12	2.01	1.15
Kind mit körperlicher Behinderung	2.17	1.29	2.18	1.31
Kind mit Verhaltensauffälligkeit	2.56	1.36	2.55	1.28
Kind mit Migrationshintergrund	2.07	1.27	2.07	1.24
Gesamt	2.17	1.26	2.18	1.22

Anmerkungen: MZP = Messzeitpunkt.

Fragestellung 2:
Können bisherige Ergebnisse
zu den Itemcharakteristika repliziert
werden?

Wie in Tabelle 2 dargestellt, zeigen die Items über alle Vignetten (Kind ohne Auffälligkeiten, Kind mit geistiger Behinderung, Kind mit Lernbehinderung, Kind mit körperlicher Behinderung, Kind mit Verhaltensauffälligkeit, Kind mit Migrationshintergrund) eine mittlere Inter-Item-Korrelation ($\alpha = .48$ bis $\alpha = .73$). Die Items weisen eine mittlere bis hohe Trennschärfe ($r_{it} = .48$ bis $r_{it} = .68$) auf. Bei Betrachtung der Gesamtskala zeigt sich eine gute Skalenhomogenität ($\alpha = .91$).

Fragestellung 3:
Können bisherige Ergebnisse
zur konvergenten Validität repliziert
werden?

Die Kurzversion der CATCH-Skala (Schwab, 2015a) diene zur konvergenten Validierung durch ein Außenkriterium. Tabelle 3 stellt diese Korrelation für die einzelnen Kindvignetten getrennt und für die Gesamtskalen dar. Die diskriminanten Korrelationen der unterschiedlichen Fallvignetten sind in allen Fällen kleiner als die konvergenten Korrelationen. Die Gesamtskala des SoDiS weist eine mittlere bis hohe Korrelation mit der CATCH-Skala auf ($r = .73, p < .01$). Diese Werte deuten auf eine gute konvergente Validität hin.

Tabelle 2: Inter-Item-Korrelationen und Trennschärfen

Item	Item				r_{it}
	Klassenraum	Schulhof	Kinosaal	Essensraum	
Klassenraum	1.0				.68
Schulhof	.48**	1.0			.54
Kinosaal	.49**	.48**	1.0		.52
Essensraum	.73**	.48**	.51**	1.0	.48

Anmerkungen: ** $p < .01$.

Tabelle 3: Konvergente Validität: Korrelation des SoDiS und der CATCH-Skala (Schwab, 2015a)

SoDiS	CATCH-Skala					
	GB	LB	KB	ADHS	MI	Gesamtskala
GB	.72**	.41**	.47**	.33**	.40**	.59**
LB	.42**	.64**	.43	.38**	.28**	.53**
KB	.49**	.45**	.65**	.39**	.33**	.59**
ADHS	.45**	.43**	.41**	.67**	.27**	.58**
MI	.30**	.27**	.25**	.18**	.66**	.41**
Gesamtskala	.58**	.53**	.52**	.57**	.47**	.73**

Anmerkungen: GB = Kind mit geistiger Behinderung, LB = Kind mit Lernbehinderung, KB = Kind mit Körperbehinderung, ADHS = Kind mit einer ADHS-Symptomatik, MI = Kind mit Migrationshintergrund
 ** $p < .01$.

**Fragestellung 4:
Weisen die Items des SoDiS
eine akzeptable bis gute
Test-Retest-Reliabilität auf?**

Die Test-Retest-Reliabilitätswerte wurden für die einzelnen Behinderungsformen und das Kind mit Migrationshintergrund getrennt ausgewertet. Tabelle 4 zeigt die Korrelationen der Testwerte zwischen dem ersten und dem zweiten Messzeitpunkt. Diese liegen im mittleren Bereich. Den niedrigsten Zusammenhang zeigt die Vignette des Kindes ohne Auffälligkeiten (OA) ($r = .60$, $p < .01$), der höchste Zusammenhang der Messwerte zeichnet sich beim Kind mit einer geistigen Behinderung ab ($r = .72$, $p < .01$). Die Gesamtskala des SoDiS weist eine gute Test-Retest-Reliabilität von $r = .83$ ($p = .01$) auf.

Diskussion und Limitationen

Die vorliegende Studie schließt an eine vorhergehende Untersuchung zur Prüfung der Testgüte des SoDiS an (Gerullis & Huber, 2016). Zentral waren eine Replikation der Ergebnisse zur konvergenten Validität anhand eines Außenkriteriums und die Prüfung der Test-Retest-Reliabilität der Skala.

Im Folgenden werden diese beiden zentralen Themenkomplexe diskutiert.

Ausprägung der Werte

Zur Schaffung eines Überblicks über die Ausprägung der Daten und somit der sozialen Distanz konnte in Anlehnung an die Ergebnisse der Vorgängerstudie (Gerullis & Huber, 2016) dargestellt werden, dass die Distanz zu allen Kindern mit Auffälligkeiten (Kind mit geistiger Behinderung, Kind mit Lernbehinderung, Kind mit körperlicher Behinderung, Kind mit Verhaltensauffälligkeit, Kind mit Migrationshintergrund) signifikant höher war als die Distanz zum Kind ohne Auffälligkeiten. Das Verfahren zeigte in diesem Zusammenhang keine Boden- oder Deckeneffekte. Diese Ergebnisse betreffen beide Messzeitpunkte. In der Varianzanalyse zeigte sich kein Haupteffekt für den Messzeitpunkt, aber für die Fallvignette. Ausschlaggebend für das Antwortverhalten ist somit die Art der Beeinträchtigung des Kindes in der jeweiligen Fallvignette. Nach Bonferronikorrektur fanden sich hier signifikante Unterschiede zwischen Kindern ohne Auffälligkeiten und der Vignette der Kinder mit Verhaltensauffälligkeit und geistiger Behinderung, sowie zwischen dem Kind mit Lernbehinderung und dem Kind mit Verhaltensauffälligkeit. Dieser Befund

Tabelle 4: Test-Retest-Reliabilität des SoDiS: Korrelationen zwischen Messzeitpunkt (MZP)1 und 2.

MZP 1	MZP 2						
	OA	GB	LB	KB	ADHS	MI	Gesamtskala
OA	.60**						
GB	.30**	.72**					
LB	.36**	.45**	.69**				
KB	.26**	.46**	.41**	.70**			
ADHS	.26**	.41**	.38**	.41**	.68**		
MI	.26**	.38**	.34**	.36**	.26**	.71**	
Gesamtskala	.37**	.58**	.53**	.59**	.51**	.50**	.83*

Anmerkungen: MZP = Messzeitpunkt, OA = Kind ohne Auffälligkeiten, GB = Kind mit geistiger Behinderung, LB = Kind mit Lernbehinderung, KB = Kind mit Körperbehinderung, ADHS = Kind mit einer ADHS-Symptomatik, MI = Kind mit Migrationshintergrund; * $p < .05$, ** $p < .01$.

zur differenzierten Wahrnehmung von Kindern mit Lern- und Verhaltensauffälligkeiten deckt sich mit Ergebnissen von Gerullis & Huber (2016) und Gasser (2012).

Itemcharakteristika und konvergente Validität

Die Items zeigen eine hohe Skalenhomogenität. In dieser Arbeit konnte mit $\alpha = .91$ eine deutlich höhere Homogenität gezeigt werden als in der Vorgängerstudie ($\alpha = .68$). Grund dafür könnten die höheren Inter-Item-Korrelationen sein. Dieses Ergebnis der Homogenität zeigt, dass die Items ein zusammenhängendes Konstrukt erfassen. Auch die Trennschärfen zeigten in der Replikationsstudie deutlich höhere Werte ($r_{it} = .481$ bis $r_{it} = .684$ im Vergleich zu $r_{it} = .25$ bis $r_{it} = .49$). Gründe dafür könnten die heterogenere Stichprobe dieser Replikationsstudie (in Bezug auf das Alter der Teilnehmerinnen und Teilnehmer und das Vorkommen von Kindern mit sonderpädagogischem Förderbedarf) und die damit verbundene größere Varianz in den Antworten sein (Bühner, 2011). Wir können davon ausgehen, dass der SoDiS gut zwischen Schulkindern mit hoher und niedriger Merkmalsausprägung in der sozialen Distanz trennt. Im Bereich der konvergenten Validität konnte anhand einer Korrelation ($r = .73, p < .01$) zu einer Kurzversion der CATCH-Skala (Schwab, 2015a) gezeigt werden, dass die Items das intendierte Konstrukt akzeptabel bis gut abbilden. Dieses Ergebnis repliziert gemäß unserer Hypothese die Aussagen der Vorgängerstudie ($r = .72, p = .01$) (Gerullis & Huber, 2016) und zeigt, dass die Items das Konstrukt valide erfassen (Bühner, 2011).

Test-Retest-Reliabilität

Explizite Maße zur Betrachtung von Einstellungen haben in bisherigen Studien relativ hohe Reliabilitätswerte aufgezeigt (Huskinson & Haddock, 2004). Wendet man den Blick jedoch auf implizite Verfahren – wie den hier untersuchten SoDiS – zeigten sich

bislang maximal akzeptable interne Konsistenzen und Test-Retest-Reliabilitätswerte (Cunningham, Preacher & Banaji, 2001). Amelang und Schmidt-Atzert (2006) nennen drei Einflussmaße auf die Reliabilität eines Tests: den Kontext, die Itemmenge und das Test-Retest-Intervall. So unterscheiden sie in der Reliabilitätsanalyse zwischen leistungsbezogenen Erhebungen (Speedtests) und Persönlichkeitsfragebögen, die stabile Merkmale erfassen. Wir ordnen den SoDiS eher Letzterem zu und gehen demnach von einer stabilen Ausprägung der sozialen Distanz aus. So zeigt sich die Angemessenheit der in dieser Studie verwendeten Methode der Bestimmung der Test-Retest-Reliabilität: Um die Stabilität der Antworten untersuchen zu können, umfasste diese Studie zwei MZP. Im Rahmen der Analyse zeigte sich eine akzeptable bis gute Stabilität der Messung ($r = .83, p = .01$). Betrachten wir die geringe Itemmenge im SoDiS und das relativ große Test-Retest-Intervall von vier Wochen, die eher hinderlich in der Erreichung hoher Test-Retest-Korrelationen sein könnten (Amelang & Schmidt-Atzert, 2006), so können wir von einer stabilen Merkmalsausprägung ausgehen. In Bezug auf das Kind ohne Auffälligkeiten zeigt sich eine niedrigere Stabilität der Messung ($r = .60, p < .01$). Dieses könnte durch die wenig markante Information über das Kind erklärt werden. Wir gehen davon aus, dass die Teilnehmerinnen und Teilnehmer im Vergleich zu den anderen dargestellten Kindern hierbei keine stabilen Anhaltspunkte zur Bewertung dieser Vignette haben und eher willkürlich antworten. Für die anderen Vignetten ergaben sich höhere Korrelationen zwischen MZP 1 und MZP 2 von $r = .68$ bis $r = .72$ ($p = .00$). Die Gesamtskala zeigte in Bezug auf die von Bühner (2011) dargestellten Grenzwerte eine mittlere Korrelation von $r = .83$ ($p < .01$). Unsere Hypothese der akzeptablen Test-Retest-Reliabilität des SoDiS kann somit bestätigt werden.

Limitationen und Ausblick

Der SoDiS zeigt im Blickwinkel der klassischen Testtheorie eine gute Testgüte. Weitere Itemanalysen zur Untersuchung der Itemstruktur zur empirischen Begründung der Itemskalierungen sollten jedoch im Rahmen der Item-Response-Theorie anhand von Rasch-Modellen vollzogen werden (siehe Bühner, 2011). Es sollten weitere Vignetten mit Kindern mit unterschiedlichen Auffälligkeiten hinzugezogen werden, die im inklusiven Kontext bedeutend werden. So könnten auch Kinder mit einem Förderbedarf im Bereich Sehen oder Hören mit einbezogen werden, für die es im SoDiS bisher keine Vignetten gibt. Der SoDiS zeigt somit nicht nur gute inhaltliche Anknüpfungspunkte, sondern auch eine, im Vergleich zu anderen impliziten Verfahren, gute Zuverlässigkeit. Insgesamt erfassen die Items valide und reliabel die soziale Distanz von Grundschülerinnen und Grundschulern zu Kindern mit besonderen Auffälligkeiten und stellen einen guten Ansatz zur Ergründung sozialer Aspekte inklusiver schulischer Prozesse dar.

Literatur

- About, F.E., & Mitchell, F.G. (1977). Ethnic role taking: The effects of preferences and self-identification. *International Journal of Psychology, 12*, 1-17.
- Amelang, M., & Schmidt-Atzert, L. (2006). *Psychologische Diagnostik und Intervention*. (4. Aufl.). Heidelberg: Springer.
- Bakker, J.T.A., & Bosman, A.M.T. (2003). Self-image and peer acceptance of Dutch students in regular and special education. *Learning Disability Quarterly, 26*, 5-14.
- Bless, G., & Mohr, K. (2007). Die Effekte von Sonderunterricht und gemeinsamen Unterricht auf die Entwicklung von Kindern mit Lernbehinderungen. In: J. Borchert, H. Goetze, J. Walter & F.B. Wember, (Hrsg.), *Sonderpädagogik des Lernens* (337-383). Göttingen: Hogrefe.
- Bogardus, E.S. (1933). *A Social Distance Scale*. Verfügbar unter: https://www.brocku.ca/MeadProject/Bogardus/Bogardus_1933.html [14.05.2016].
- Bossaert, G, Colpin, H., Pijl, S. J., & Petry, K. (2012). Loneliness among students with special educational needs in mainstream seventh grade. *Research in Developmental Disabilities, 33*, 1888-1897.
- Bossaert, G., & Petry, K. (2013). Factorial validity of the Chedoke-McMaster attitudes towards children with handicaps scale (CATCH). *Research in Developmental Disabilities, 34*, 1336-1345.
- Bukowski, W. M., Cillessen, A. H. N. & Velásquez, A. M. (2012). Peer Ratings. In B. P. Laursen, T. D. Little, & N. A. Card (Hrsg.), *Handbook of developmental research methods* (211–228). New York: Guilford Press.
- Bühner, Markus. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. aktualisierte und erweiterte Auflage). München u.a: Pearson Studium.
- Catapusan, B. T. (1954). Social distance in the Philippines. *Sociology & Social Research, 38*, 309-312.
- Cillessen, A. H. N. (2011). Sociometric Methods. In K. H. Rubin, W. M. Bukowski, & B. P. Laursen (Hrsg.), *Handbook of peer interactions, relationships, and groups*. New York: Guilford.
- Clack, B.T. (2007). *Racial contact and isolation in a post-apartheid South African school*. Unpublished doctoral dissertation, University of Lancaster, UK.
- Connolly, P., Fitzpatrick, S., Gallagher, T., & Harris, P. (2006). Addressing diversity and inclusion in the early years in conflict-affected societies: A case study of the Media Initiative for Children, Northern Ireland. *International Journal of Early Years Education, 14*, 263-278.
- Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: consistency, stability, and convergent vali-

- dity. *Psychological Science*, 12(2), 163-170.
- De Boer, A., Pijl, S.J. & Minnaert, A. (2012). Students Attitudes towards Peers with Disabilities: A review of the literature. *International Journal of Disability, Development and Education*, 59, 379-392.
- Feyerer, E. (1998). *Behindern Behinderte? Integrativer Unterricht in der Sekundarstufe I*. Innsbruck/Wien: Studienverlag.
- Frederickson, N.L., & Funham, A.F. (2001). The long-term stability of sociometric status classification: A longitudinal study of included pupils who have moderate learning difficulties and their mainstream peers. *Journal of Child Psychology and Psychiatry*, 42, 581-592.
- Gasser, L, Chilver-Stainer, J., Buholzer, A., & Perrig-Chiello, P. (2012). Soziales und moralisches Denken von Kindern über den Ein- und Ausschluss behinderter Kinder. *Zeitschrift für Pädagogische Psychologie*, 26, 31-42.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464-1480.
- Gerullis, A. & Huber, C. (2016). Soziale Distanz in inklusiven Settings (SoDiS) – Entwicklung und erste Validierung von Items zur Messung sozialer Distanz von Grundschulkindern zu Kindern mit Behinderungen. *Empirische Sonderpädagogik* (4), 327-345.
- Gerullis, A. & Huber, C. (2017). I agree to sit next to you. Does that mean I like you? Measuring using the wrong tapeline – the lack of ‘Social Distance’ measurements for inclusive school development and research. *Journal of Research in Special Educational Needs*. doi: 10.1111/1471-3802.12399.
- Haerberlin, U. (1991) die Integration von leistungsschwachen Schülern. Ein Überblick über empirische Forschungsergebnisse zu Wirkungen von Regelklassen, Integrationsklassen und Sonderklassen auf „Lern-behinderte“. *Zeitschrift für Pädagogik*, 37, 167-189.
- Holzkamp, K. (1964). Über soziale Distanz. *Psychologische Beiträge*, 7, 558-581.
- Huber, C. (2006). Soziale Integration in der Schule?!: Eine empirische Untersuchung zur sozialen Integration von Schülern mit sonderpädagogischem Förderbedarf im gemeinsamen Unterricht. Marburg: Tec-tum.
- Huber, C. (2008). Jenseits des Modellversuchs: Soziale Integration von Schülern mit sonderpädagogischem Förderbedarf im Gemeinsamen Unterricht – Eine Evaluationsstudie. *Heilpädagogische Forschung*, 34, 2-14.
- Huber, C., & Wilbert, J. (2012). Soziale Ausgrenzung von Schülern mit sonderpädagogischem Förderbedarf und niedrigen Schulleistungen im gemeinsamen Unterricht. *Empirische Sonderpädagogik*, 2, 147-165.
- Huskinson, L. H., & Haddock, G. (2004). Individual differences in attitude structure and the accessibility of the affective and cognitive components of attitude. *Social Cognition*, 24(4), 453-468.
- Kavale, K. A., & Forness, S. R. (1996). Social Skill deficits and learning disabilities. A meta-analysis. *Journal of Learning Disabilities*, 29, 226-337.
- Koster, M., Nakken, H., Pijl, S. J. & van Houten, E. (2009). Being part of the peer group: a literature study focusing on the social dimension of inclusion in education. *International Journal of Inclusive Education*, 13(2), 117-140.
- Lauber, C., Nordt, C., Falcanto, L., & Rössler, W. (2004). Factors influencing social distance toward people with mental illnesses. *Community Mental Health Journal*, 40, 265-273.
- Laws, G., & Kelly, E. (2005). The attitudes and friendship intentions of children in United Kingdom mainstream schools towards peers with physical or intellectual disabilities. *International Journal of Disability, Development & Education*, 52, 79-99.

- Lindsay, G. (2007). Educational psychology and the effectiveness of inclusive education/mainstreaming. *British Journal of Educational Psychology*, 77, 1-24.
- Lindemann, H. (2016). Konstruktion und empirische Validierung eines Instrumentes zur Erfassung der Einstellungen von Schülerinnen und Schülern gegenüber Peers mit Beeinträchtigung, anderer Herkunft und niedrigem sozioökonomischen Status. *Empirische Sonderpädagogik*, 1, 5-21.
- Lüke, T. & Grosche M. (2018). Konstruktion und Validierung der Professionsunabhängigen Einstellungsskala zum Inklusiven Schulsystem (PREIS). *Empirische Sonderpädagogik*, 1, 3-20.
- Moreno, J.L. (1974). *Die Grundlagen der Soziometrie*. Wege zur Neuordnung der Gesellschaft. Opladen: Leske & Budrich.
- Myklebust, J.O. (2002). Inclusion or exclusion? Transitions among special needs students in upper secondary education in Norway. *European Journal of Special Needs Education*, 17, 251-263.
- Nowicki, E. A. (2006). Across-sectional multivariate analysis of children's attitudes towards disabilities. *Journal of Intellectual Disability Research*, 50, 335-348.
- Preuss-Lausitz, U. (2009). Integrationsforschung. Ansätze, Ergebnisse und Perspektiven. In: Eberwein, H. (Hrsg.): *Handbuch der Integrationspädagogik. Kinder mit und ohne Beeinträchtigung lernen gemeinsam*. 7. Auflage (S. 458-470). Weinheim/Basel: Beltz.
- Rosenbaum, P.L., Armstrong, R.W., & King, S.M. (1985). Childrens attitudes toward disabled peers: A self-report measure. *Journal of Pediatric Psychology*, 11, 517-530.
- Ruijs, N., Peetsma, T., & Van der Veen, I. (2010). The presence of several students with special educational needs in inclusive education and the functioning of students with special educational needs. *Educational Review*, 62, 1-37.
- Schwab, S. (2015a). Einflussfaktoren auf die Einstellung von SchülerInnen gegenüber Peers mit unterschiedlichen Behinderungen. In *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 47, 1-11.
- Schwab, S. (2015b). Lehrersicht der sozialen Partizipation von Grundschulern: Ergebnisse einer Studie mit dem Lehrer Fragebogen zur Erfassung der sozialen Partizipation. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 84, 234-245.
- Verna, G. (1981). Use of a free-response task to measure children's race preferences. *The Journal of Genetic Psychology*, 138, 87-93.
- Wocken (1992). Bewältigung von Andersartigkeit. Untersuchungen zur Sozialen Distanz in verschiedenen Schulen. In: Gehrmann P. & Hüwe, B. (1993). *Forschungsprofile der Integration von Behinderten*, 1992. Bochumer Symposion. Essen: Neue Deutsche Schule Verlagsgesellschaft.
- Wocken, H. (1999). Schulleistungen in heterogenen Lerngruppen. In: Eberwein, H. (Hrsg.): *Integrationspädagogik. Kinder mit und ohne Behinderung lernen gemeinsam*. (315-320). Weinheim/Basel: Beltz.

Anita Gerullis

Bergische Universität Wuppertal
 Rehabilitationswissenschaften mit dem
 Förderschwerpunkt emotionale und
 soziale Entwicklung
 Gaußstraße 20
 42119 Wuppertal
 E-Mail: gerullis@uni-wuppertal.de

Erstmalig eingereicht: 26.09.2018

Überarbeitung eingereicht: 04.12.2018

Angenommen: 12.03.2019