

## Preprint

### Stufen der sozialen Partizipation nach Einschätzung von Regel- und Integrationslehrkräften

Susanne Schwab & Markus Gebhardt

Schwab, S. & Gebhardt, M. (2016). Stufen der sozialen Partizipation nach Einschätzung von Regel- und Integrationslehrkräften. *Empirische Pädagogik*, 30(1), 43–66.

#### Zusammenfassung

Die vorliegende Studie untersucht die soziale Partizipation von 886 Grund- und Sekundarstufenschülern auf Basis der Einschätzung von Klassen- und Integrationslehrern. Ziel ist es, den Fragebogen zur Erfassung der sozialen Partizipation (SPQ), welcher bislang lediglich nach klassischer Testtheorie analysiert wurde, mittels des Rasch-Modells zu prüfen und zu kürzen. Auf Basis des Gesamtscores des SPQs wird ein Stufenmodell der sozialen Partizipation entwickelt. Die Ergebnisse zeigen, dass nach dem Ausschluss einzelner Items das Rasch Modell als gültig angenommen werden kann. Die soziale Partizipation kann anhand der folgenden drei Stufen interpretiert werden: Außenseiter (Stufe 1), Akzeptanz (Stufe 2) und Freundschaft (Stufe 3). Die Übereinstimmung der beiden Lehrerurteile korreliert insgesamt zu  $r=.43$ . In 71% der Fälle können die Schüler anhand der Einschätzungen der Klassen- und Integrationslehrkräfte der identischen Stufe zugewiesen werden. Bei 13% der Fälle sind die Lehrkräfte genau gegenteiliger Meinung. Daraus kann geschlossen werden, dass die Klassen- und die Integrationslehrkraft je einen eigenen Blick auf bestimmte Schüler hat.

Schlüsselbegriffe: soziale Partizipation, Inklusion, sonderpädagogischer Förderbedarf, Lehrereinschätzung;

#### Levels of social participation in the opinion of control and integration teachers

##### Abstract

This study examines the social participation of 886 primary and secondary students based on the assessments by regular and special needs teachers. To this extent, the social participation questionnaire (SPQ) was analyzed using probabilistic test theory (Rasch model) and, subsequently, a multi-stage model of social participation was developed based on the total score. The results show, that after eliminating some items, the Rasch model can be accepted as valid and social participation can be divided into the following three levels: outsiders (level 1), acceptance (level 2) and friendship (level 3). The correlation indicating the agreement between the teacher ratings is .43. In 71% of the cases, students are assigned the same level by both teachers. However, in 13% of the cases, the teachers'

opinions are completely opposite. This indicates that the two teachers each have their own particular views on some of the students.

Keywords: social participation, inclusion, special educational needs, teacher assessment;

## **Einleitung**

Im Zuge steigender schulischer Inklusion im europäischen Raum sieht es die Forschung als zentrale Aufgabe, die Effektivität von gemeinsamem Unterricht von Schülern mit sonderpädagogischem Förderbedarf (SPF) und ihren Peers ohne SPF aufzuzeigen und Problemlagen zu identifizieren. Während in den Köpfen der Bevölkerung die schulische Förderung von Schulkindern mit SPF als eher schwierig eingeschätzt wird und die soziale Partizipation von Schülern mit SPF hingegen positiver gesehen wird (siehe z.B. Schwab, 2014), zeigen empirische Begleitforschungen eher ein konträres Bild. Hinsichtlich schulischer Leistungsentwicklung sprechen die Literaturbefunde mehrheitlich für positive Effekte, sowohl für Schüler mit SPF als auch für ihre Mitschüler ohne SPF (siehe dazu z.B. Dessemontet & Bless, 2013; Dessemontet, Bless & Morin, 2012; Kojac, Kuhl, Kroth, Pant, & Stanat, 2014). Betreffend der sozialen Partizipation wird im folgenden Abschnitt genauer ausgeführt, dass die Ergebnisse bisheriger Forschungen (z.B. Bossaert, de Boer, Frostad, Pijl & Petry, 2015; Koster, Pijl, Nakken & van Houten, 2010; Schwab, im Druck) mehrheitlich ein problematisches Bild für Schüler mit SPF zeigen.

### *Soziale Partizipation von Schülern mit und ohne SPF*

Die Operationalisierung der sozialen Partizipation kann über unterschiedliche Kernbereiche erfolgen: soziale Beziehungen (Freundschaften bzw. Teil eines sozialen Netzwerks sein), soziale Interaktionen (z.B. gemeinsam verbrachte Freizeit, oder gemeinsames Arbeiten an Aufgaben, Teilnahme an Gruppenaktivitäten), soziale Akzeptanz (z.B. soziale Position) sowie Selbstwahrnehmung (z.B. Einsamkeitsgefühl, Schulzufriedenheit, Selbsteinschätzung der Peerakzeptanz) (siehe dazu Koster, Nakken, Pijl & van Houten, 2009a sowie Bossaert, Colpin, Pijl & Petry, 2013a). Dabei weichen die Ergebnisse je nach Operationalisierung etwas voneinander ab. Grütter, Meyer und Glenz (2015) zeigen sogar, dass innerhalb einer ähnlichen Operationalisierung (Verwendung der identischen soziometrischen Frage, aber Auswertung von unterschiedlichen soziometrischen Maßen: Freundschaften (wechselseitige Nennungen), Beliebtheit (Anzahl der Nennungen anderer, Zugehörigkeit und Zentralität in Schülercliquen) unterschiedliche Ergebnisse zu erwarten sind. Die Autoren kamen zu dem Ergebnis, dass sich Schüler mit und ohne SPF zwar nicht in der Anzahl der Freundschaften, jedoch in der Beliebtheit und der Zugehörigkeit in Schülercliquen von ihren Peers ohne SPF unterscheiden. Selbst bei der Verwendung identischer Messinstrumente bzw. Maße sind internationale Vergleichsstudien schwierig, da sowohl die uneinheitlichen Definitionen des SPFs als auch die Unterschiede in den Schulsystemen Vergleiche erschweren. Bossaert et al. (2015) kommen bei dem Vergleich der sozialen Partizipation von Schülern aus Belgien (two-track system), Holland

(multi-track system) und Norwegen (one-track system) zu dem Schluss, dass in dem inklusiveren Schulsystem (in diesem Fall in Norwegen) wie auch in den beiden Vergleichsländern die soziale Akzeptanz der Schüler mit SPF signifikant geringer ist als jene von den Peers ohne SPF. Jedoch merken die Autoren kritisch an, dass internationale Vergleiche aufgrund der unterschiedlichen Definitionen von Schülern mit SPF schwierig sind. Beispielsweise enthält das norwegische Sample mehr Schüler mit schwerwiegenden Beeinträchtigungen und geistiger Behinderung. Dennoch werden hier zusammenfassend jene internationalen Studien dargestellt, welche sich nicht lediglich auf einen Kernbereich, sondern auf alle vier genannten Kernbereiche (soziale Beziehungen, soziale Interaktionen, soziale Akzeptanz, Selbstwahrnehmung der sozialen Partizipation) erstrecken. Diesbezüglich ist anzumerken, dass nur wenige Studien alle diese Teilbereiche untersuchten.

Koster et al. (2010) führten eine Studie mit holländischen Grundschulern durch. Dabei fanden die Autoren keine signifikanten Gruppenunterschiede zwischen Schülern mit und ohne SPF in der Selbsteinschätzung der sozialen Partizipation. In den anderen drei Bereichen (Interaktionen erfasst durch Beobachtungen; Peerrating für die Peerakzeptanz und soziometrische Nominierungen für Freundschaft) fanden sich hingegen Unterschiede zugunsten der Schüler ohne SPF, wobei die Effektstärken im moderaten Bereich liegen. Spannend an der Untersuchung ist zudem, dass Schüler ohne SPF weniger in Interaktionen mit ihren Peers mit SPF treten, Lehrkräfte treten jedoch im Vergleich zu Schülern ohne SPF mit Schüler mit SPF häufiger in Interaktionen. Für österreichische Grund- und Sekundarstufenschüler untersuchte Schwab (2015a) die soziale Partizipation. Dabei zeigte sich für alle vier Subbereiche (Interaktionen mit Peers in der Pause, erfasst mittels soziometrischer Ratingmethode; wechselseitige Freundschaften und soziale Präferenz (in-degree), erfasst über soziometrische Nominierungen und Selbsteinschätzung der sozialen Integration) eine signifikant geringere soziale Partizipation für Schüler mit SPF. Die Effektstärke für die Gruppenunterschiede zwischen Schülern mit und ohne SPF zeigt sich dabei für die Interaktionen, die Freundschaften und die Peerakzeptanz hoch, für die Selbsteinschätzung hingegen ist lediglich ein moderater Effekt zu verzeichnen. Bossaert, Colpin, Pijl und Petry (2013b) untersuchten Sekundarstufenschüler in Belgien (Flanders) und kamen auch zu dem Ergebnis, dass Schüler mit SPF (Autismus-Spektrum-Störung; ASS) in allen vier Bereichen (Interaktionen mit Peers in der Pause, erfasst mittels soziometrischer Ratingmethode; wechselseitige Freundschaften und soziale Präferenz (in-degree) und das soziale Selbstkonzept) über eine geringere soziale Partizipation verfügen. Zudem verglichen die Autoren die soziale Partizipation von den Schülern mit ASS mit einer Subgruppe von Schülern mit motorischen und/oder sensorischen Störungen und fanden diesbezüglich keine Gruppenunterschiede.

Neben den genannten Studien wurden zudem Fremdeinschätzungen über Lehrkräfte zur Beurteilung der sozialen Partizipation eingesetzt. Hier ist grundsätzlich festzustellen, dass es lediglich einen Fragebogen gibt, den „Social Participation Questionnaire“ (SPQ), welcher alle vier Teilbereiche der sozialen Partizipation erfasst. Dieses Instrument wurde von Koster et al. (siehe Koster, Nakken, Pijl, van Houten & Spelberg, 2008; sowie Koster, Timmerman, Nakken, Pijl & van Houten, 2009b)

erstellt und mit Grundschullehrkräften erprobt. Bossaert, Martens, Vanmarsenille, Vertessen & Petry (2013c) prüften die Güte des Lehrereinschätzfragebogens zudem für SchülerInnen der Sekundarstufe. Anschließend übersetzten Schwab und Hessels das Instrument ins deutschsprachige und führten erste Analysen zu den üblichen Gütekriterien (Reliabilität, Trennschärfe, Faktorenstruktur) für Grund- und Sekundarstufenschüler durch (siehe dazu Schwab, 2015b; Hessels & Schwab, eingereicht). Die genannten Befragungen der Lehrkräfte bestätigen allesamt, dass die soziale Partizipation von Schülern mit SPF im Vergleich zu ihren Peers, welche dieselben Schulklassen besuchen, aber keinen (festgestellten) SPF haben, geringer eingeschätzt wird. Allerdings ist für alle berichteten Studien kritisch anzumerken, dass nicht geprüft wurde, ob das verwendete Messinstrument zur Messung der sozialen Partizipation gegenüber den Gruppen (Schüler mit und ohne SPF) auch Messinvarianz aufweist. Somit ist unklar, ob der Fragebogen wirklich verwendet werden darf, um Unterschiede zwischen Schülern mit und ohne SPF zu erfassen.

Ein weiteres Problem, welches aktuell in der Forschung zur sozialen Partizipation vorliegt, ist, dass nicht klar definiert wurde, was unter einer hohen bzw. einer niedrigen sozialen Partizipation zu verstehen ist. Zwar werden bei den Mittelwertvergleichen immer wieder niedrigere Mittelwerte für Schüler mit SPF berichtet, jedoch findet weder eine theoretische Einordnung noch eine empirische Überprüfung der Skalierung statt. Der SPQ wurde somit als reines Forschungsinstrument entwickelt, ohne dass eine Normierung stattfand. In den genannten Studien zur Lehrkräfteeinschätzung mittels SPQ wurden die vier Teilbereiche der sozialen Partizipation gleichwertig nebeneinandergestellt und zu einem Summenwert addiert. Vorher wurden die Dimensionen mittels einer explorativen Faktoranalyse geprüft. Für die Bildung eines Summenwertes für die soziale Partizipation wäre der Nachweis der Gültigkeit eines ein-parametrischen Item-Response Modells von Vorteil. Damit ist es erlaubt, den Summenwert der Skala als intervallskalierte Variable und als eindimensionalen Wert der sozialen Partizipation zu betrachten (Rost, 2004).

Um die soziale Partizipation als eindimensionales, ein-parametrischen Item-Response Modell zu konstruieren, ist es notwendig, die Dimension anhand der Schwierigkeit der Itemparameter zu definieren. Unseres Erachtens können die Dimensionen des SPQs als Stufen betrachtet werden. Auf der geringsten Stufe der sozialen Partizipation ist ein aktives negatives Verhalten gegenüber des betroffenen Schülers (z.B. Ausgrenzung) einzuordnen. Dagegen ist eine gemeinsame Akzeptanz als Klassenmitglied eine höhere Stufe. Die höchste Stufe ist eine wechselseitige Freundschaft mit anderen Mitschülern.

### **Zielsetzung**

Lehrereinschätzungen sind Beurteilungen über Schüler, welche für die zukünftige schulische Karriere wichtig sind. Anhand der Einschätzung der Schüler reflektieren die Lehrkräfte auch ihre eigene Meinung bezogen auf den Schüler. Dies ist insbesondere wichtig für Lehrerfort- und -ausbildung,

welche in Bezug auf soziale Partizipation notwendig ist. Lehrerratings sollten einfach anwendbar und leicht auszuwerten sein. Daher sollte aus den Items einfach den Summenwert benutzt werden dürfen. Testtheoretisch ist dies nur bei einparametrischen Modellen möglich (Rost, 2004). Das Ziel dieser Studie ist es somit, den SPQ als Lehrerrating zur Erfassung sozialer Partizipation von Schülern hinsichtlich seiner Güte nach prohabilitischer Testtheorie zu prüfen und eine kürzere Version des SPQ bestehend aus allen vier Dimensionen als eindimensionales, einparametrisches Modell zu skalieren. Die Skala muss zwischen Schülerinnen und Schülern mit und ohne SPF messinvariant sein, damit das Lehrerrating auch in inklusiven Klassen eingesetzt werden kann. Hierbei ist es notwendig, dass die Überprüfung der Messinvarianz auf Itemebene vorliegt, da aufgrund des SPFs anzunehmen ist, dass gewisse Aspekte der sozialen Partizipation als subgruppenspezifisch gesehen werden. Des Weiteren ist es wichtig, dass Klassenlehrkräfte und Integrationslehrkräfte die Ratingskala ähnlich beantworten.

## **Untersuchungsmethode**

### ***Stichprobe***

Die Daten für die vorliegenden Berechnungen stammen aus dem Projekt ATIS-SI (siehe Schwab, 2015c), wobei lediglich die Daten der Lehrereinschätzung zum zweiten Messzeitpunkt (Ende des Schuljahres) verwendet wurden. Von den insgesamt 1047 teilnehmenden Schülern (aus 61 Schulklassen) besuchten 564 Schüler eine Integrationsklasse. Auf der Lehrerebene liegen über insgesamt 794 Schüler (75.8%) Einschätzungen vom Klassenlehrer und über 405 Schüler (71.8%) Einschätzungen der Integrationslehrkraft vor. Über 313 Schüler (=55.5% der 564 Integrationsklassenschüler) liegen Bewertungen beider Lehrkräfte vor. Über 92 Schüler liegt nur eine Einschätzung der Integrationslehrkraft vor und über 481 Schüler nur eine Einschätzung der Regellehrkraft. Im Datensatz befinden sich somit 886 Schüler, 112 mit einem diagnostizierten SPF, welcher sich zumeist auf den Bereich Lernen bezieht.

### ***Erhebungsinstrument***

Als Erhebungsinstrument wurde die von Bossaert et al. (2013c) verwendete Version des SPQ mit insgesamt 24 Items (siehe Tabelle 1) verwendet (siehe dazu auch Schwab, 2015a sowie Hessels & Schwab, eingereicht). Der Fragebogen besteht aus vier Dimensionen: Interaktionen (9 Items), Akzeptanz (5 Items), Freundschaft (5 Items) und Selbstsicht (5 Items), welche alle auf einer fünfstufigen Likertskala mit den Kategorien 1 = „trifft überhaupt nicht zu“, 2 = „trifft eher nicht zu“, 3 = „weder noch“ 4 = „trifft eher zu“, 5 = „trifft völlig zu“ zu beantworten sind. Die interne Konsistenz für die Originalskalen weisen je nach Subskala und Lehrkraft für die vorliegende Stichprobe ein Cronbach's Alpha von mindestens .76 auf. Für einen Wert in der sozialen Partizipation werden alle Items miteinander summiert.

## Statistische Analysen

Alle statistischen Analysen wurden in den Statistikpaketen R (R Core Team, 2014) und SPSS durchgeführt. Für die Berechnungen des Raschmodells wurde das R Paket „pairwise“ (Heine, 2014) verwendet. Dieses Paket wird zur expliziten Berechnung der Itemparameter im Raschmodell nach der Methode des paarweisen Itemvergleichs (Choppin, 1968, 1985) verwendet. Diese Methode eignet sich insbesondere zur Bestimmung der stichprobeninvarianten Itemparameter für die Kalibrierung eines gegebenen Itempools (Choppin, 1968). Der Pairwise-Schätzer wird gewählt, da dieser auch bei fehlenden Daten und kleinen Stichproben robust schätzt (Heine & Tarnai, 2015).

## Ergebnisse

### Deskriptive Kennwerte

In Tabelle 1 sind die einzelnen Mittelwerte und Standardabweichungen der Einschätzung über die 24 Items angeführt. Insgesamt zeigte sich, dass alle Schüler sehr positiv bewertet wurden und nur wenige Schüler niedrige Einschätzungen in den Items erhielten. Der Modalwert aller Items liegt jeweils in der höchsten Antwortkategorie. Demzufolge ist die vorliegende Verteilung der Items jeweils rechtsgipflig. Dies ist bei Einschätzungsskalen des sozialen Verhaltens und der sozialen Partizipation häufig der Fall (Schwab, 2014).

Tabelle 1: Übersicht über die deutsche Version der SPQ Items mit Mittelwerten und Standardabweichungen getrennt für Klassen- und Integrationslehrer

| Nr                   | Item des SPQ   | Klassenlehrer |                | Integrationslehrer |                |
|----------------------|--|---------------|----------------|--------------------|----------------|
|                      |  | N             | M<br>(SD)      | N                  | M<br>(SD)      |
| <b>Interaktionen</b> |  |               |                |                    |                |
| 1                    | Das Kind hat offensichtlich Spaß mit seinen/ihren Mitschülern (gemeinsam lachen).  | 794           | 4.38<br>(0.78) | 405                | 4.52<br>(0.73) |
| 3*                   | Die Mitschüler schließen das Kind generell von Aktivitäten aus.  | 793           | 4.40<br>(0.92) | 405                | 4.57<br>(0.86) |
| 4*                   | Die Mitschüler warten auf das Kind nach der Schule, jagen es, beschimpfen es, provozieren es und ähnliches.                                | 793           | 4.59<br>(0.79) | 405                | 4.75<br>(0.66) |
| 5                    | Die Mitschüler fragen das Kind, ob es mitspielen möchte.   | 794           | 4.11<br>(0.92) | 403                | 4.23<br>(1.05) |
| 7*                   | Die Mitschüler mobben/hänseln das Kind.  | 793           | 4.51<br>(0.83) | 404                | 4.67<br>(0.79) |
| 10                   | Das Kind nimmt an Spielen teil ohne dass der Lehrer die anderen Kinder dazu auffordern muss.   | 793           | 4.28<br>(0.91) | 403                | 4.51<br>(0.85) |
| 11*                  | Die Mitschüler lachen das Kind aus.  | 792           | 4.48<br>(0.88) | 404                | 4.59<br>(0.85) |
| 16                   | Wenn die Schüler in einem Kreis sitzen, setzen sich andere Schüler freiwillig neben das Kind.  | 792           | 4.36<br>(0.83) | 404                | 4.51<br>(0.85) |
| 19                   | Die anderen Kinder arbeiten spontan mit dem Kind, ohne dass dies vom Lehrer initiiert werden muss.   | 792           | 4.07<br>(1.09) | 402                | 4.31<br>(1.04) |
| <b>Akzeptanz</b>     |  |               |                |                    |                |
| 2                    | Wenn es nötig ist, sind die Mitschüler gewillt die Regeln eines Spiels so anzupassen, dass das Kind mitspielen kann.                       | 782           | 4.04<br>(0.93) | 362                | 4.09<br>(0.99) |
| 12                   | Wenn gemeinsam gearbeitet wird, berücksichtigen die Kinder die Lernfähigkeit des Kindes (Was das Kind tun kann und was nicht).             | 751           | 3.90<br>(0.95) | 353                | 4.05<br>(1.00) |
| 13                   | Wenn ein Spiel ausgewählt wird, berücksichtigen die Kinder die Fähigkeiten und Fertigkeiten des Kindes (Was das Kind kann und nicht kann). | 749           | 3.90<br>(0.93) | 353                | 4.08<br>(1.00) |

|                     |  |     |                |     |                |
|---------------------|--|-----|----------------|-----|----------------|
| 15                  | Die Mitschüler ergreifen Partei für das Kind wenn Schüler anderer Klassen oder Schulen das Kind schlecht behandeln (würden).   | 794 | 4.04<br>(0.94) | 402 | 4.19<br>(0.94) |
| 23                  | Die Mitschüler sind darauf vorbereitet dem Kind zu helfen. Z.B.: Schuhe binden, Dinge aufzuheben, Material und Unterlagen zu ordnen etc. (Hilfe im positiven Sinne; nicht bemuttern o.ä.). | 728 | 3.70<br>(0.99) | 328 | 3.47<br>(1.13) |
| <b>Freundschaft</b> |  |     |                |     |                |
| 6                   | Das Kind ist Teil von einem Freundeskreis.   | 786 | 4.26<br>(0.94) | 402 | 4.40<br>(0.92) |
| 8                   | Das Kind hat einen oder mehrere gute Freunde in der Gruppe (mit dem es gemeinsame Unternehmungen macht und regelmäßigen Kontakt hat).  | 788 | 4.19<br>(0.98) | 401 | 4.37<br>(0.95) |
| 14                  | Das Kind wird von einem oder mehreren Mitschülern zum Spielen nach der Schule eingeladen.  | 748 | 3.83<br>(1.10) | 390 | 3.87<br>(1.15) |
| 17                  | Ein oder mehrere Kinder ergreifen die Initiative mit dem Kind auch während der Ferien zu spielen.  | 732 | 3.87<br>(1.04) | 379 | 3.87<br>(1.11) |
| 20                  | Das Kind wird zu Geburtstagsfeiern eingeladen.   | 718 | 3.84<br>(1.09) | 364 | 3.88<br>(1.12) |
| <b>Selbstsicht</b>  |  |     |                |     |                |
| 9                   | Das Kind fühlt sich als Teil der Gruppe und sieht sich selbst nicht als Außenseiter.   | 787 | 4.15<br>(1.00) | 399 | 4.39<br>(0.98) |
| 18*                 | Das Kind fühlt sich in der Schule einsam.  | 776 | 4.23<br>(1.04) | 395 | 4.45<br>(0.98) |
| 21                  | Das Kind darf es selbst sein in der Schule und muss sich nicht verstellen oder vorgeben anders zu sein als es ist.   | 757 | 4.28<br>(0.92) | 390 | 4.59<br>(0.67) |
| 22*                 | Das Kind fühlt sich mehr gemobbt/gehänselt als andere Kinder.  | 787 | 4.26<br>(1.08) | 400 | 4.52<br>(0.97) |
| 24                  | Das Kind geht gerne zur Schule.  | 794 | 3.97<br>(0.96) | 397 | 4.23<br>(0.84) |

\*rekodierte Items

## Ergebnisse auf Basis der Klassenlehreinschätzungen

### *Vergleich der Modelle mit fünf, drei und zwei Antwortkategorien*

Die Angaben der Klassenlehrer werden von allen Items als eine Gesamtskala zusammengenommen und mit einem Partial Credit Model berechnet. Überprüft wird dabei, ob die Antwortkategorien entlang des zu messenden Kontinuums ordinal geordnet sind (Rost, 2004). Es fällt auf, dass sich die Antwortkategorien der einzelnen Items überlagern und die Lehrkräfte somit bei den Antwortkategorien nicht genau differenzieren können. In Tabelle 2 ist dies durch die punkt-biserial Korrelation der Itemkategorien mit dem WLE-Gesamtscore der soziale Partizipation (Warm's Weighted Mean Likelihood Estimate) dargestellt. Je höher die Kategorie des Items, desto höher sollten die punkt-biserialen Korrelationskoeffizienten der einzelnen Items sein. Wenn jedoch niedrigere Kategorien von Items einen höheren Korrelationskoeffizienten haben (siehe dazu auch Tabelle 2), liegt eine Unordnung in den Kategorien vor (Linacre, 1999). Dies bedeutet, dass einzelne Kategorien der Items nicht häufig genug nach dem Modell benutzt wurden, da eine ordinale Struktur in den Kategorien angenommen wird. Auch die Thurstonian-Thresholds zeigen für beide Modelle an, dass einzelne Kategorien der Lehrer überlappen und die Schüler nicht trennscharf geschätzt werden (Wu & Adams, 2007). In Tabelle 2 ist dies für das Modell mit fünf Kategorien (links) und für das Modell mit drei Kategorien (rechts) der Fall. Auffällig ist, dass im Modell mit drei Kategorien die neutrale Antwortkategorie „weder noch“ negativ gesehen wird und nahe an der Kodierung „trifft nicht zu“ liegt. Aus diesem Grund wird für das Lehrerrating ein Modell mit zwei Kategorien gebildet, bei dem die Kategorien „trifft überhaupt nicht zu“, „trifft eher nicht zu“ und „weder noch“ als eine Kategorie 0

und „trifft eher zu“ sowie „trifft völlig zu“ als Kategorie 1 betrachtet werden. Die punkt-biseriale Korrelation für dieses Modell ist in Tabelle 3 dargestellt.

Tabelle 2: Punkt-biseriale Korrelation der Kategorien mit dem WLE-Gesamtscore (Klassenlehreereinschätzungen)

|    | Modell mit 5 Kategorien |                |                |                |                | Modell mit 3 Kategorien |                |                |
|----|-------------------------|----------------|----------------|----------------|----------------|-------------------------|----------------|----------------|
|    | Kategorie<br>1          | Kategorie<br>2 | Kategorie<br>3 | Kategorie<br>4 | Kategorie<br>5 | Kategorie<br>1          | Kategorie<br>2 | Kategorie<br>3 |
| 1  | <b>-0.09</b>            | <b>-0.25</b>   | -0.24          | -0.39          | 0.60           | -0.35                   | -0.28          | 0.45           |
| 2  | <b>-0.09</b>            | <b>-0.26</b>   | -0.35          | -0.32          | 0.62           | <b>-0.36</b>            | <b>-0.40</b>   | 0.56           |
| 3  | <b>-0.06</b>            | <b>-0.21</b>   | -0.30          | -0.30          | 0.52           | <b>-0.26</b>            | <b>-0.36</b>   | 0.46           |
| 4  | <b>-0.18</b>            | <b>-0.29</b>   | -0.34          | -0.33          | 0.74           | -0.45                   | -0.38          | 0.61           |
| 5  | <b>-0.09</b>            | <b>-0.23</b>   | -0.31          | -0.35          | 0.58           | <b>-0.32</b>            | <b>-0.36</b>   | 0.49           |
| 6  | <b>-0.17</b>            | <b>-0.24</b>   | -0.34          | -0.37          | 0.68           | <b>-0.38</b>            | <b>-0.38</b>   | 0.56           |
| 7  | <b>-0.07</b>            | <b>-0.26</b>   | -0.26          | -0.33          | 0.55           | <b>-0.36</b>            | <b>-0.29</b>   | 0.47           |
| 8  | <b>-0.15</b>            | <b>-0.23</b>   | -0.30          | -0.43          | 0.68           | <b>-0.35</b>            | <b>-0.37</b>   | 0.52           |
| 9  | <b>-0.20</b>            | <b>-0.30</b>   | -0.34          | -0.29          | 0.74           | -0.47                   | -0.39          | 0.65           |
| 10 | <b>-0.16</b>            | <b>-0.30</b>   | -0.37          | -0.27          | 0.75           | -0.44                   | -0.41          | 0.61           |
| 11 | <b>-0.13</b>            | <b>-0.23</b>   | -0.44          | -0.11          | 0.67           | <b>-0.36</b>            | <b>-0.46</b>   | 0.62           |
| 12 | <b>-0.17</b>            | <b>-0.24</b>   | -0.44          | -0.13          | 0.71           | <b>-0.37</b>            | <b>-0.47</b>   | 0.63           |
| 13 | <b>-0.18</b>            | <b>-0.27</b>   | -0.35          | -0.28          | 0.73           | <b>-0.42</b>            | <b>-0.40</b>   | 0.60           |
| 14 | <b>-0.15</b>            | <b>-0.08</b>   | -0.44          | -0.12          | 0.69           | <b>-0.21</b>            | <b>-0.48</b>   | 0.58           |
| 15 | <b>-0.16</b>            | <b>-0.27</b>   | -0.34          | -0.37          | 0.71           | -0.41                   | -0.38          | 0.58           |
| 16 | <b>-0.18</b>            | <b>-0.28</b>   | -0.33          | -0.36          | 0.72           | -0.44                   | -0.39          | 0.61           |
| 17 | <b>-0.22</b>            | <b>-0.30</b>   | -0.29          | -0.21          | 0.74           | -0.50                   | -0.35          | 0.66           |
| 18 | <b>-0.23</b>            | <b>-0.24</b>   | -0.36          | -0.18          | 0.72           | <b>-0.44</b>            | <b>-0.44</b>   | 0.67           |
| 19 | <b>-0.22</b>            | <b>-0.23</b>   | -0.38          | -0.19          | 0.74           | <b>-0.41</b>            | <b>-0.45</b>   | 0.67           |
| 20 | <b>-0.09</b>            | <b>-0.29</b>   | -0.31          | -0.37          | 0.72           | <b>-0.38</b>            | <b>-0.37</b>   | 0.56           |
| 21 | <b>-0.07</b>            | <b>-0.34</b>   | -0.32          | -0.31          | 0.67           | -0.44                   | -0.36          | 0.60           |
| 22 | <b>-0.05</b>            | <b>-0.19</b>   | -0.29          | -0.36          | 0.60           | <b>-0.24</b>            | <b>-0.36</b>   | 0.45           |
| 23 | <b>-0.16</b>            | <b>-0.27</b>   | -0.25          | -0.27          | 0.58           | -0.42                   | -0.28          | 0.53           |
| 24 | <b>-0.10</b>            | <b>-0.28</b>   | -0.25          | -0.21          | 0.60           | -0.36                   | -0.27          | 0.48           |

In Tabelle 3 sind die Itemparameter, die punkt-biseriale Korrelation und die Item Fit-Werte dargestellt. Die Itemparameter stellen die Schwierigkeiten der Items dar, welche für die Berechnung der Personenparameter benutzt werden. Die punkt-biseriale Korrelation kann bei einem Modell mit nur zwei Kategorien als Trennschärfe betrachtet werden und ist in diesem Fall zufriedenstellend.



Betrachtet man die Koeffizienten des Outfits, Infits und der punkt-biserialen Korrelation, so ist zu erkennen, dass die meisten Items in Bezug auf das Raschmodell eine gute Passung aufweisen (Wright & Masters, 1982). Die fett gedruckten Kennwerte in Tabelle 3 zeigen Items, deren Mean-Square Fit-Werte bedeutsam von ihrem Erwartungswert 1 abweichen. Liegt der Wert unter 1, so ist die Vorhersage der Daten höher als vom Raschmodell angenommen. Diese Items sind dann „zutrennscharf“ und haben eher ein Antwortmuster der Guttman Pattern. Werte, die signifikant über 1 liegen sind problematisch, da diese einen Underfit aufweisen. Erklärungen dafür sind Decken- bzw. Bodeneffekte bezüglich des ‚targeting‘ von Personenfähigkeit und Itemschwierigkeit sowie geratene/zufällige Antworten (Linacre, 2002). Wie in Tabelle 3 ersichtlich ist liegen die Infit- und Outfit-Werte hier zwischen .7 und 1.5. Somit kann angenommen werden, dass die Messung nicht negativ beeinflusst wird (Linacre, 2002; 2003), jedoch ist für eine Schülerbewertung ein strengerer Bereich (0.8-1.2) anzuwenden (Wright & Linacre, 1994). Demzufolge werden jene Items, die nicht innerhalb dieses Bereichs liegen, ausgeschlossen. Darüber hinaus sollten die vier Dimensionen in dem Gesamtscore einen gleich starken Einfluss haben, insofern werden drei Items pro Dimension für die weitere Skala ausgewählt (siehe Tabelle 4).

Tabelle 3: Itemkennwerte im Modell mit 2 Antwortkategorien (Klassenlehreereinschätzungen)

| Itemnr. | Itemparameter | Punkt-<br>biseriale<br>Korrelation | Outfit       | Infit        |
|---------|---------------|------------------------------------|--------------|--------------|
| 1       | -1.28         | 0.47                               | 1.06         | 1.09         |
| 2       | -0.48         | 0.62                               | 0.88         | 0.88         |
| 3       | -0.92         | 0.50                               | 1.03         | 1.10         |
| 4       | -0.21         | 0.66                               | <b>0.80</b>  | <b>0.82*</b> |
| 5       | -1.12         | 0.56                               | 0.91         | 0.92         |
| 6       | -0.53         | 0.60                               | 0.85         | 0.94         |
| 7       | -1.02         | 0.51                               | 0.98         | 1.01         |
| 8       | -1.24         | 0.58                               | 0.72         | 0.91         |
| 9       | 0.27          | 0.69                               | <b>0.80</b>  | <b>0.82*</b> |
| 10      | 0.44          | 0.66                               | 0.92         | 0.98         |
| 11      | 1.05          | 0.66                               | 1.06         | 1.04         |
| 12      | 1.08          | 0.67                               | 1.06         | 1.01         |
| 13      | 0.20          | 0.65                               | 0.94         | 0.96         |
| 14      | 1.70          | 0.59                               | <b>1.49*</b> | <b>1.22*</b> |
| 15      | -0.46         | 0.62                               | 0.86         | 0.89         |
| 16      | -0.34         | 0.65                               | <b>0.79</b>  | <b>0.87*</b> |
| 17      | 0.93          | 0.67                               | 0.98         | 0.98         |

|    |       |      |              |              |
|----|-------|------|--------------|--------------|
| 18 | 1.01  | 0.68 | 1.00         | 0.97         |
| 19 | 1.17  | 0.67 | 1.06         | 1.00         |
| 20 | -0.17 | 0.59 | 1.06         | 1.00         |
| 21 | 0.01  | 0.64 | 0.88         | 0.91         |
| 22 | -0.44 | 0.48 | <b>1.20</b>  | <b>1.26*</b> |
| 23 | -0.05 | 0.55 | 1.16         | 1.11         |
| 24 | 0.38  | 0.50 | <b>1.50*</b> | <b>1.32*</b> |

\* signifikanter Z-Wert,  $p < .01$ ;

Tabelle 4: Itemkennwerte nach Itemselektion im Modell mit zwei Antwortkategorien (Klassenlehreereinschätzungen)

| Itemnummer | Itemparameter | Punkt-biseriale<br>Korrelation | Outfit | Infit |
|------------|---------------|--------------------------------|--------|-------|
| 1          | -1.50         | 0.52                           | 1.00   | 1.08  |
| 2          | -0.61         | 0.64                           | 0.92   | 0.94  |
| 5          | -1.25         | 0.57                           | 0.97   | 1.00  |
| 10         | 0.33          | 0.67                           | 0.99   | 1.03  |
| 11         | 0.96          | 0.63                           | 1.20*  | 1.15* |
| 13         | 0.07          | 0.67                           | 0.95   | 0.99  |
| 17         | 0.76          | 0.73                           | 0.94   | 0.90* |
| 18         | 0.86          | 0.74                           | 0.92   | 0.88* |
| 19         | 1.04          | 0.73                           | 0.99   | 0.92  |
| 20         | -0.33         | 0.61                           | 1.05   | 1.04  |
| 21         | -0.16         | 0.67                           | 0.92   | 0.95  |
| 23         | -0.18         | 0.57                           | 1.15   | 1.13* |

\* signifikanter Z-Wert,  $p < .01$ ;

### ***Prüfung der Messinvarianz***

Eine wichtige Grundlage für die Verwendung des gekürzten Modells mit zwei Antwortkategorien ist, dass diese subgruppeninvariant für Geschlecht und Schüler mit und ohne SPF misst. In Abbildung 1 sind die Itemparameter der beiden Subgruppen (der Klassenlehreereinschätzung) auf der X- und Y-Achse abgetragen. Diese verlaufen entlang der Winkelhalbierenden, wobei die 95% Konfidenzintervalle der Itemparameterpunktschätzer als Ellipsen dargestellt sind. Wie in der Abbildung 1 zu erkennen ist, legt der Verlauf der Itemparameter einerseits die Modellgeltung nahe und belegt andererseits, dass die Items in beiden Gruppen (annähernd) gleiche Schwierigkeiten aufweisen. Alle Items sind nahe der Winkelhalbierenden. Im Vergleich der Schüler mit und ohne SPF

ist das Item 17 problematisch, da es für Schüler mit SPF im Vergleich zu den regulären Schülern schwieriger ist.

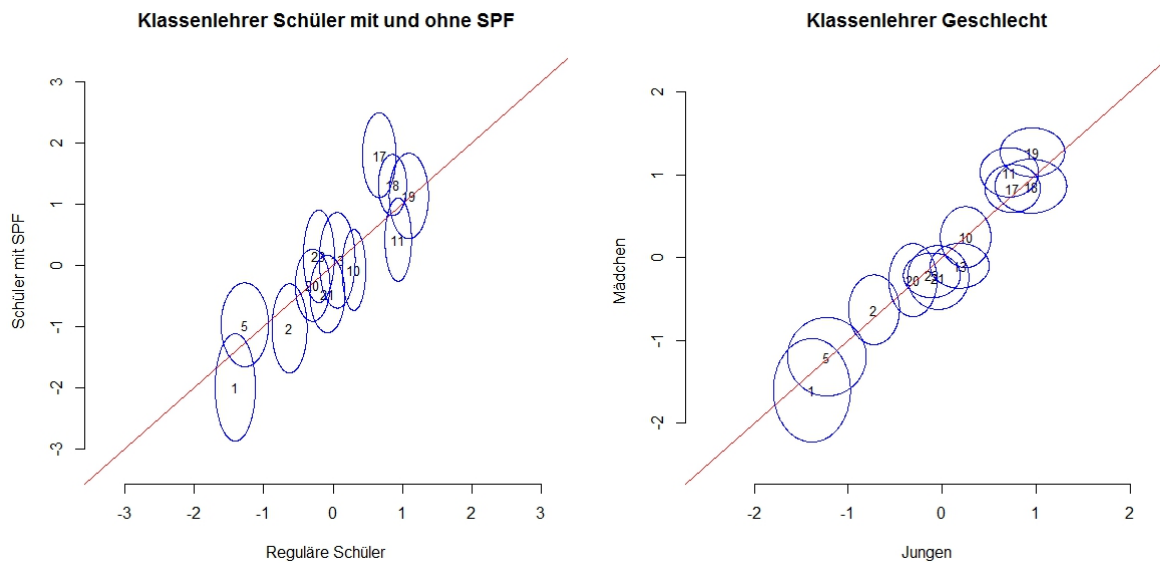


Abbildung 1. Graphischer Modelltest der Schülerinnen und Schüler mit und ohne SPF und dem Geschlecht der Schülerinnen und Schüler in der Klassenlehreinschätzung

### Validierung der Ergebnisse durch die Einschätzung der Integrationslehrer

Um zu überprüfen, ob die Integrationslehrkräfte die Schüler mit derselben Skala einschätzen können, wurde für die Ratings der Integrationslehrer ebenfalls ein gekürztes Modell mit zwei Antwortkategorien gerechnet. In Tabelle 5 ist zu erkennen, dass die Skala auch bei den Integrationslehrkräften den Wertebereich von 0.8 bis 1.2 der Fit-Werte einhält und das Modell die Daten zufriedenstellend wiedergibt.

Tabelle 5: Itemkennwerte nach Itemselektion im Modell mit zwei Antwortkategorien (Integrationslehreinschätzungen)

| Item | Itemparameter | Punkt-biseriale Korrelation | Outfit | Infit |
|------|---------------|-----------------------------|--------|-------|
| 1    | -1.61         | 0.50                        | 0.88   | 1.07  |
| 2    | -0.83         | 0.62                        | 0.87   | 0.94  |
| 5    | -1.36         | 0.59                        | 0.82   | 0.92  |
| 10   | 0.67          | 0.59                        | 1.20   | 1.19* |
| 11   | 0.99          | 0.61                        | 1.16   | 1.15* |
| 13   | 0.26          | 0.66                        | 0.96   | 0.94  |

|    |       |      |      |       |
|----|-------|------|------|-------|
| 17 | 1.11  | 0.73 | 0.95 | 0.86* |
| 18 | 1.13  | 0.71 | 1.03 | 0.93  |
| 19 | 1.07  | 0.71 | 1.01 | 0.93  |
| 20 | -0.49 | 0.61 | 0.94 | 0.92  |
| 21 | -0.46 | 0.49 | 1.24 | 1.15  |
| 23 | -0.48 | 0.59 | 0.94 | 1.00  |

\* signifikanter Z-Wert,  $p < .01$ ;

Der grafische Modelltest zeigt eine Subgruppeninvarianz bei den Items 10 und 11 im Vergleich der Schülerinnen und Schülern mit und ohne SPF auf. An die regulären Schülerinnen und Schüler werden hier höhere Anforderungen gestellt als an Schülerinnen und Schüler mit SPF.

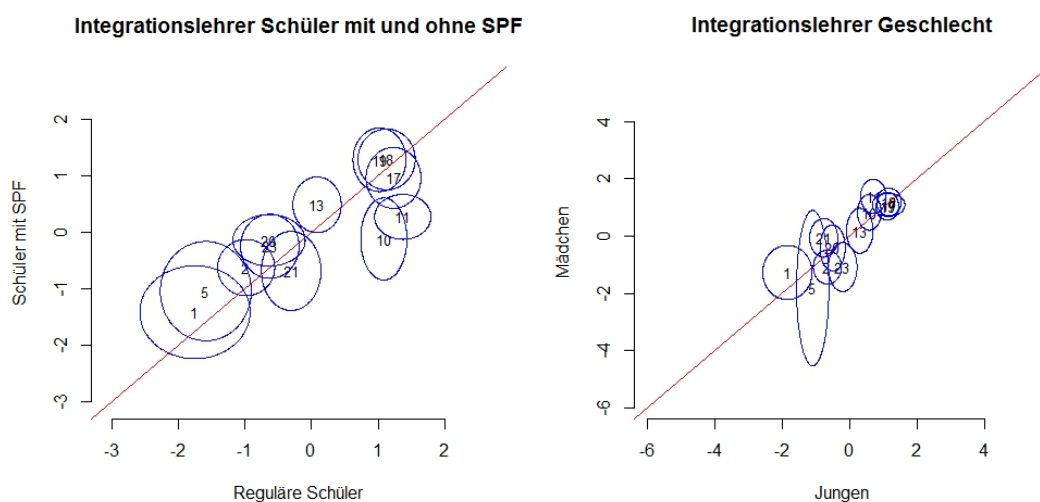


Abbildung 2. Graphischer Modelltest der Schülerinnen und Schüler mit und ohne SPF und dem Geschlecht der Schülerinnen und Schüler in der Integrationslehreinschätzung

### Item-Personen-Map bei Klassen- und Integrationslehrern

Die Items 10, 11 und 17 wurden aufgrund von Subgruppeninvarianz für Schüler mit SPF für die weiteren Analysen ausgeschlossen. Nachfolgend sind die Raschmodelle der Klassenlehrkräfte (links) und der Integrationslehrkräfte (rechts) in Abbildung 3 durch die Person-Item-Map gegenübergestellt. In der Person-Item-Map sind die Personenparameter als vertikales Histogramm mit den Itemparametern der einzelnen Items verbunden dargestellt. Man erkennt in beiden Bildern, dass die größte Gruppe der Schüler den höchsten Skalenscore aufweist. Die Items 18 und 19 der Dimension Freundschaft sind die schwierigsten Items, während die Items zur Interaktion und zur Selbstsicht die leichtesten Items sind. Nur wenige Schüler werden in diesem niedrigen Personenparameterbereich eingeschätzt. Anhand der Verteilung der Personen und der Items kann man Stufen der sozialen Partizipation bilden. Personen mit einem Score unter  $-0.10$  (nähe Item 2) haben eine sehr niedrige

soziale Partizipation und werden als Außenseiter (Stufe 1) eingeschätzt. Personen mit einem Wert zwischen -0.09 und 0.7 (nähe Item 13) liegen im Bereich der Akzeptanz (Stufe 2), erreichen aber keine Freundschaftsitems (Stufe 3). Dieser Bereich fängt erst ab dem Wert 0.71 an.

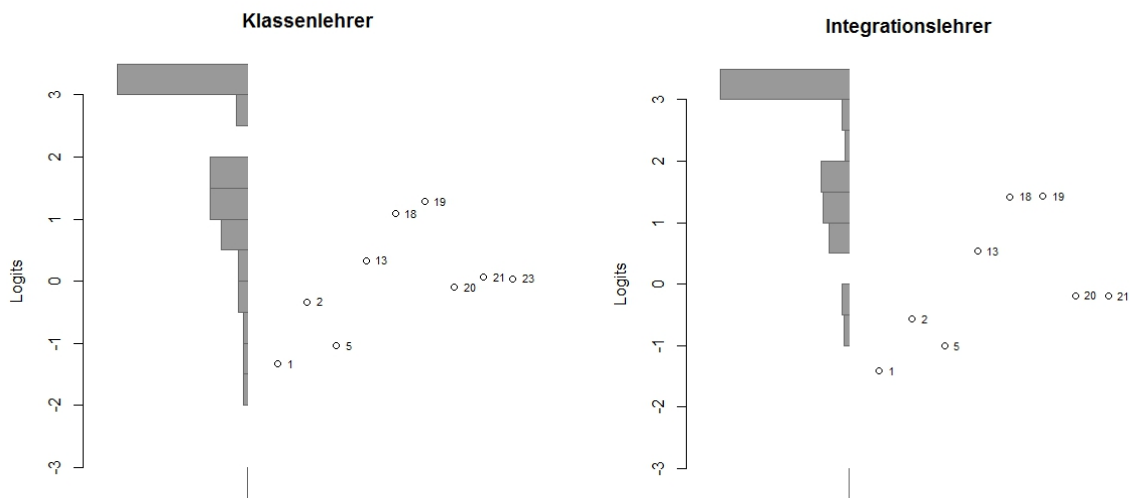


Abbildung 3: Item-Personen-Map der Schülereinschätzung der Klassen- und Integrationslehrer. Im Diagramm ist die Verteilung der Personenparameter als Histogramm den Itemparametern der Items gegenübergestellt.

Die WLE-Personenparameter zwischen Klassenlehrkräften und Integrationslehrkräften korreliert zu  $r=.43$ . Teilt man die Stichprobe nach Schülerinnen und Schülern mit und ohne SPF auf, ergibt sich bei den Schülerinnen und Schülern mit SPF eine Korrelation der Einschätzung von  $r=.45$  ( $N=74$ ) und bei Schülerinnen und Schülern ohne SPF eine Korrelation von  $r=.36$  ( $N=238$ ).

### ***Auftretenshäufigkeiten der einzelnen Stufen der sozialen Partizipation***

In der folgenden Tabelle sind die Auftretenshäufigkeiten der einzelnen Stufen der sozialen Partizipation dargestellt. Sowohl für Schüler mit als auch für Schüler ohne SPF wird die Stufe 3 von beiden Lehrern am häufigsten vergeben. Dabei ist jedoch auch erkennbar, dass die Wahrscheinlichkeit in Stufe 3 zu sein für Schüler ohne SPF deutlich höher ist (81-85%) als für Schüler mit SPF (60-61%). Umgekehrt fallen Schüler mit SPF 2-3mal häufiger in die problematische Stufe 1 (23-29%) als Schüler ohne SPF (5-14%). Zudem lässt die Tabelle erkennen, dass der Integrationslehrer die soziale Partizipation durchwegs positiver beurteilt als der Klassenlehrer (dies steht auch mit Tabelle 7 im Einklang).

Tabelle 6: Auftretenshäufigkeiten der einzelnen Stufen, getrennt für Schüler mit und ohne SPF

|  | Schüler ohne SPF | Schüler mit SPF |
|--|------------------|-----------------|
|--|------------------|-----------------|

|                    | Stufe 1       | Stufe 2       | Stufe 3        | Stufe 1       | Stufe 2       | Stufe 3       |
|--------------------|---------------|---------------|----------------|---------------|---------------|---------------|
| Klassenlehrer      | 96<br>(13.9%) | 33<br>(4.8%)  | 561<br>(81.3%) | 24<br>(28.6%) | 9<br>(10.7%)  | 51<br>(60.7%) |
| Integrationslehrer | 14<br>(4.7%)  | 32<br>(10.6%) | 255<br>(84.7%) | 23<br>(22.5%) | 18<br>(17.6%) | 61<br>(59.8%) |

Wie in Tabelle 7 zu erkennen ist, stimmt die Einstufung der Schüler aus der Sicht beider Lehrer bei 70.9% der Fälle überein. In 12.8% der Fälle liegen die Lehrkräfte in ihrer Einschätzung jedoch konträr um zwei Stufen auseinander und in 16.3% um eine Stufe.

Tabelle 7: Anzahl der Übereinstimmung zwischen Klassen- und Integrationslehrer in Bezug auf die Einstufung der Kinder und Prozentangaben der Gesamtanzahl

|               |         | Integrationslehrer |                 |                    |             |
|---------------|---------|--------------------|-----------------|--------------------|-------------|
|               |         | Stufe 1            | Stufe 2         | Stufe 3            | Alle        |
| Klassenlehrer | Stufe 1 | <b>19 (6.1%)</b>   | 9 (2.9%)        | 30 (9.6%)          | 58 (18.5%)  |
|               | Stufe 2 | 2 (0.6%)           | <b>7 (2.2%)</b> | 12 (3.8%)          | 21 (6.7%)   |
|               | Stufe 3 | 10 (3.2%)          | 28 (8.9%)       | <b>196 (62.9%)</b> | 234 (74.8%) |
|               | Alle    | 31 (9.9%)          | 44 (14.1%)      | 238 (76%)          | 313 (100%)  |

Anmerkung: Gleiche Einstufungen wurden fett markiert

## Diskussion

Ziel der vorliegenden Studie war es, den Lehrerfragebogen zur Einschätzung der sozialen Partizipation (SPQ) für den Einsatz in inklusiven Klassen zu überprüfen. Da die Skala nach der klassischen Testtheorie entwickelt und erprobt wurde (siehe dazu Koster et al., 2008; 2009b; Bossaert et al., 2013c; Schwab, 2015,a; Hessels & Schwab, eingereicht), mussten die Kategorien zusammengelegt und einzelne Items ausgeschlossen werden, um zufriedenstellende Fit-Werte zu erhalten. Es zeigt sich, dass das Modell mit zwei Antwortkategorien messvariant in Bezug auf die Einschätzung von Schülerinnen und Schülern mit und ohne SPF ist. Ein direkter Vergleich der Schülerpopulationen miteinander ist mit der ursprünglichen Skala bei den Items 10, 11 und 17 nicht sinnvoll. Bei dem Item 10 „Wenn es nötig ist, sind die Mitschüler gewillt die Regeln eines Spiels so anzupassen, dass das Kind mitspielen kann“ werden beispielsweise an Schülerinnen und Schüler ohne SPF höhere Anforderungen gestellt als an Schülerinnen und Schüler mit SPF. Die vorliegende Kurzsкала sollte in weiteren Studien überprüft werden, wobei eine Ergänzung mit Items mit höherer Itemschwierigkeit sinnvoll wäre.

Des Weiteren ist es mit der Einteilung in Stufen einfacher, den Gesamtscore inhaltlich zu verorten und zu interpretieren. Durch das Raschmodell können die einzelnen Items als Stufen der sozialen Partizipation betrachtet werden. Ein hoher oder niedriger Skalenwert kann somit als Stufe Außenseiter (Stufe 1), Akzeptanz (Stufe 2) oder Freundschaft (Stufe 3) interpretiert werden. Die deskriptiven Ergebnisse der Studie zeigen, dass ein Großteil der Schüler ohne SPF, aber auch die Mehrheit der Schüler mit SPF in die Stufe 3 eingeteilt wurden und somit sozial vollwertig in der Klasse partizipieren. Dies lässt darauf schließen, dass konform mit früheren Studien (Koster, Pijl, Nakken & van Houten, 2010) die Mehrheit aller Schüler eine zufriedenstellende soziale Partizipation aufweist. Dennoch ist das Risiko Außenseiter zu sein für Schüler mit SPF im Vergleich zu ihren Peers hoch und auch der Anteil an Schülern mit SPF, welche nicht akzeptiert werden, ist deutlich höher. Auch dieses Ergebnis geht mit bisherigen Literaturbefunden einher (siehe z.B. Huber & Wilbert, 2012; Humphrey & Symes, 2010). In der vorliegenden Studie waren diesbezüglich die Einschätzungen der Lehrkräfte ähnlich wie die Befunde bei den Untersuchungen von Schülerinnen und Schülern. Die Messinstrumente dieser Studien haben generell das Problem, hohe Deckeneffekte aufzuweisen, da Schüler identifiziert werden sollen, welche eine mangelnde soziale Partizipation aufweisen. Dies hat zur Folge, dass das Instrument nur im unteren Bereich, also im Risikobereich, genau misst, während es Personen mit „guter“ oder „ausgezeichnete“ sozialer Partizipation nicht unterscheidet. Um das gesamte Spektrum der sozialen Partizipation darzustellen wäre es sinnvoll, zusätzliche Items mit einer höheren Schwierigkeit in der Stufe Freundschaft zu konstruieren oder eine weitere Stufe Beliebtheit in der Klasse zu bilden. Denn durch den Deckeneffekt könnte die niedrige Korrelation der Lehrerübereinstimmung beeinflusst sein.

Der Vergleich der Lehrereinschätzungen von Klassen- und Integrationslehrer bezüglich der sozialen Partizipation ist ein neues Element in dieser Studie. Erstaunlich ist, dass bei Schülern ohne SPF die Lehrer auch zu unterschiedlichen Urteilen kommen. Bei über 70% der Schüler schätzen beide Lehrer eine identische Einstufung. Jedoch liegen bei mehr als einem von zehn Schülern die Urteile der beiden Lehrkräfte um zwei Stufen auseinander. Inhaltlich heißt dies, dass eine Lehrkraft der Meinung ist, dass der Schüler Außenseiter ist, wohingegen der andere Lehrer dem Schüler eine hohe Stufe der sozialen Partizipation (Freundschaft) zuspricht. Dies bedeutet, dass gewisse Schüler von den beiden Lehrkräften unterschiedlich in ihrer sozialen Partizipation wahrgenommen werden. Eine offene Frage ist, warum dies der Fall ist. Dies sollte in weiteren Studien geklärt werden.

Für eine Weiterentwicklung des Instrumentes ist es außerdem notwendig zu fragen, ob an alle Schüler der gleiche Maßstab bzgl. des Verhaltens gestellt wird und ob somit der Bezugsrahmen zur Einschätzung der sozialen Partizipation der Schüler derselbe sein kann. Da inklusiver Unterricht individualisiert erfolgt und demnach auch individuelle Verhaltensziele nach sich ziehen kann, ist es fraglich, ob alle Lehrkräfte eine ähnliche Definition von sozialer Partizipation für alle Schülerinnen und Schüler haben. So wurden in den Analysen zur Messinvarianz zwischen Schülern mit und ohne SPF bei drei Items höhere Anforderungen an die soziale Partizipation von Schülern ohne SPF

gemessen. Ziel der Instrumentenentwicklung sind jedoch Items, welche für alle Lehrkräfte und alle Schüler einen einheitlichen Referenzrahmen vorgeben, um vergleichbare Werte abzubilden.

Ebenso ist es eine Einschränkung der vorliegenden Studie, dass nicht explizit auf kulturelle Invarianz geprüft wurde. Um dies zu prüfen, müssten entweder länderübergreifende Studien oder alternativ Vergleiche anhand bestehender Daten von Kollegen (z.B. Bossaert et al., 2013c) vorgenommen werden, um zu messen, ob sich zwischen Schülern unterschiedlicher Länder Messvarianz zeigt.

Aufgrund der Bedeutung der sozialen Partizipation für die weitere psychosoziale als auch die Leistungsentwicklung von Schülern ist es wichtig, dass Lehrer die soziale Partizipation der Schüler erkennen. Insofern gilt es, in der Zukunft den Blick der Lehrkräfte mit Hilfe funktionierender Instrumente auch in diesem Bereich zu schärfen.

## Literatur

- Bossaert, G., Colpin, H., Pijl, S. J., & Petry, K. (2013a). Truly included? A literature study focusing on the social dimension of inclusion in education. *International Journal of Inclusive Education*, 17, 60-79.
- Bossaert, G., Colpin, H., Pijl, S. J., & Petry, K. (2013b). Social Participation of Students with Special Educational Needs in Mainstream Seventh Grade. *Procedia: Social and Behavioral Sciences*, 93, 1952-1956.
- Bossaert, G., de Boer, A., Frostad, P., Pijl, S.J. & Petry, K. (2015). Social participation of students with special educational needs in different educational systems. *Irish Educational Studies*, 34, 43-54.
- Bossaert, G., Martens, S., Vanmarsenille, C., Vertessen, N. & Petry, K. (2013c). De betrouwbaarheid en discriminante validiteit van se Social Participation Questionnaire in het Vlaamse onderwijs. *Pedagogische Studiën*, 90, 2-16.
- Choppin, B. (1968). Item Bank using Sample-free Calibration. *Nature*, 219, 870-872.
- Choppin, B. (1985). A fully conditional estimation procedure for Rasch model parameters. *Evaluation in Education*, 9, 29-42.
- Dessementet, R. S. & Bless, G. (2013). The impact of including children with intellectual disability in general education classrooms on the academic achievement of their low-, average-, and high-achieving peers. *Journal of Intellectual & Developmental Disability*, 38, 23-30.
- Dessementet, R. S., Bless, G. & Morin, D. (2012). Effects of inclusion on the academic achievement and adaptive behaviour of children with intellectual disabilities. *Journal of Intellectual Disability Research*, 56, 579-587.
- Grütter, J., Meyer, B. & Glenz, A. (2015). Sozialer Ausschluss in Integrationsklassen: Ansichtssache? *Psychologie in Erziehung und Unterricht*, 62, 65-82.



- Heine, J.-H. (2014). *pairwise: Rasch Model Parameters by Pairwise Algorithm*. R package version 0.2.5. Verfügbar unter: <http://CRAN.R-project.org/package=pairwise> [28.10.2014]
- Heine, J. H. & Tarnai, Ch. (2015). Pairwise Rasch Model Item Parameter Recovery under Sparse Data Conditions. *Psychological Test and Assessment Modeling*, 57(1),3-36.
- Hessels, M. G. P. & Schwab, S. (eingereicht). Sociale participatie in het secundair onderwijs in Oostenrijk gemeten met de Social Participation Questionnaire.
- Huber, C. & Wilbert, J. (2012). Soziale Ausgrenzung von Schülern mit sonderpädagogischem Förderbedarf und niedrigen Schulleistungen im gemeinsamen Unterricht. *Empirische Sonderpädagogik*, 2, 147-165.
- Humphrey, N. & Symes, W. (2010). Responses to bullying and use of social support among pupils with autism spectrum disorders (ASDs) in mainstream schools: a qualitative study. *Journal of Research in Special Educational Needs*, 10, 82-90.
- Kocaj, A., Kuhl, P., Kroth, A. J., Pant, H. A. & Stanat, P. (2014). Wo lernen Kinder mit sonderpädagogischem Förderbedarf besser? Ein Vergleich schulischer Kompetenzen zwischen Regel- und Förderschulen in der Primarstufe. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66, 165-191.
- Koster, M., Nakken, H., Pijl, S.J., van Houten, E. & Spelberg, H.C.L. (2008). Assessing social participation of pupils with special needs in inclusive education: the construction of a teacher questionnaire. *Educational Research and Evaluation: An International Journal on Theory and Practice*, 14, 395-409.
- Koster, M., Nakken, H., Pijl, S. J. & van Houten, E. J. (2009a). Being part of the peer group: A literature study focussing on the social dimension of inclusion in education. *International Journal of Inclusive Education*, 13, 117-140.
- Koster, M., Pijl, S. J., Nakken, H. & van Houten, E. (2010). Social participation of students with special needs in regular primary education in The Netherlands. *International Journal of Disability, Development and Education*, 57, 59-75.
- Koster, M., Timmerman, M.E., Nakken, H., Pijl, S.P. & van Houten, E. J. (2009b). Evaluating social participation of pupils with special needs in regular primary schools. Examination of a teacher questionnaire. *European Journal of Psychological Assessment*, 25, 213-222.
- Linacre, J.M. (1999). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3:2, 103-122.
- Linacre, J.M. (2002). What do Infit and Outfit, Mean-square and Standardized mean? *Rasch Measurement Transactions*, 16, 878.
- Linacre, J.M. (2003). Rasch Power Analysis: Size vs. Significance: Standardized Chi-Square Fit Statistic. *Rasch Measurement Transactions*, 17, 918.
- Rost, J. (2004). *Lehrbuch Testtheorie - Testkonstruktion*. Bern: Huber.

- Schwab, S. (2014). Schulische Förderung und soziale Integration von Schülern mit sonderpädagogischem Förderbedarf in Integrationsklassen - Eine explorative Analyse in der Steiermark. *Empirische Pädagogik*, 28(3), 259-274.
- Schwab, S. (2015a). Lehrersicht der sozialen Partizipation von Grundschulern – Ergebnisse einer Studie mit dem Lehrerfragebogen zur Erfassung der sozialen Partizipation. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*.
- Schwab, S. (2015b). Social dimensions of inclusion in education of 4th and 7th grade pupils in inclusive and regular classes: outcomes from Austria. *Research in Developmental Disabilities*.
- Schwab, S. (201c). Einstellung zur Integration im Zusammenhang mit sozialer Inklusion – Eine Fragebogenerhebung in österreichischen Integrations- und Regelschulklassen. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 84, 66-67.
- Schwab, S. (im Druck, a). Partizipation. In I. Hedderich, G. Biewer, J. Hollenweger & R. Markowetz (Hrsg.), *Handbuch Inklusion und Sonderpädagogik*.
- Wright, B. D., & Masters, G. N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Wright, B. D. & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, 8, 370.
- Wu, M. & Adams, R. (2007). *Applying the Rasch Model to Psycho-Social Measurment. A Practical Approach*. Melbourne: Educational Measurement Solutions.