



Effekte des Good Behavior Game bei Grundschüler_innen mit externalisierenden Verhaltensproblemen

Tobias Hagen¹ , Jannik Nitz¹ , Fabienne Brack¹, Dennis Christian Hövel² 
und Thomas Hennemann¹

¹Department Heilpädagogik und Rehabilitation, Universität zu Köln, Deutschland

²Interkantonale Hochschule für Heilpädagogik (HfH), Zürich, Schweiz

Zusammenfassung: *Hintergrund:* Das Good Behavior Game (GBG; Barrish, Saunders & Wolf, 1969) ist eine wirksame verhaltensmodifikatorische Intervention zur Verbesserung des Verhaltens von Schüler_innen im Unterricht. Allerdings ist der Kenntnisstand zur Effektivität der Maßnahme bei Schüler_innen mit externalisierenden Verhaltensproblemen auf individueller Ebene insgesamt gering. *Methoden:* Die aktuelle Studie untersucht die Auswirkungen des GBG auf das störende Verhalten (SV) und das lernförderliche Verhalten (LV) von 20 Grundschüler_innen mit externalisierenden Verhaltensproblemen der dritten Klasse auf individueller Ebene. Es wurde ein Multipler-Grundraten-Versuchsplan mit A-B Design verwendet. Zur fallübergreifenden Schätzung der Effekte wurden die Daten mittels hierarchischer stückweiser Regression analysiert. *Ergebnisse:* Die Befunde der Mehrebenenanalysen unter Berücksichtigung aller 20 Einzelfälle weisen sowohl für die Spielphase (GBG-Phase) als auch für den gesamten Schultag signifikante Level-Effekte auf beiden Dimensionen (SV und LV) nach. Während sich das lernförderliche Verhalten verbessert, geht das Störverhalten zurück. *Diskussion:* Demnach zeigt sich ein unmittelbarer Interventionseffekt auf das Verhalten der Schüler_innen. Die Ergebnisse bestätigen bisherige Untersuchungen zur Wirksamkeit des GBG bei Schüler_innen mit Verhaltensproblemen. Eine wesentliche Erweiterung stellen die Einzelfallbefunde zur Verhaltensentwicklung der Schüler_innen außerhalb der Spielphasen dar (Transfereffekte).

Schlüsselwörter: Good Behavior Game, Unterrichtsstörungen, lernförderliches Verhalten, Einzelfallstudie, Mehrebenenanalyse

Effects of the Good Behavior Game on Primary School Students with Externalizing Behavior Problems

Abstract: *Background:* The Good Behavior Game (GBG; Barrish, Saunders & Wolf, 1969) is an effective behavior modification intervention for improving students' classroom behavior. However, less is known about the effects on individual students with externalizing behavior problems. *Methods:* The current study examined the impact of the GBG on disruptive behavior (DB) and academic engagement (AE) of 20 at-risk 3rd grade elementary students at the individual level. A multiple baseline design with an A-B format was used. Hierarchical Piecewise Regression was applied to analyze the data across cases. *Results:* The findings of the multilevel approach revealed significant level effects on both dimensions (DB and AE) for the play phase (GBG phase) as well as for the whole school day. While academic engagement improves, disruptive behavior decreases. *Discussion:* Thus, there is an immediate intervention effect on students' behavior. Our results support previous research indicating that the GBG is an appropriate classroom-based intervention for students with behavior problems and expand the findings regarding students' behavioral development outside the GBG phase based on single-case data.

Keywords: Good Behavior Game, disruptive behavior, academic engagement, single-case study, hierarchical piecewise regression

Einleitung

Schulische Maßnahmen zur Verbesserung des Lern- und Arbeitsverhaltens, zur Stärkung der sozialen Integration sowie zur Reduktion von störendem Verhalten (z. B. in die Klasse rufen oder nicht auf dem Platz sitzen bleiben) weisen eine große Relevanz für die Praxis auf, was unter ande-

rem die Häufigkeit von Verhaltensproblemen unter Kindern und Jugendlichen belegt. Die KiGGS-Studie (Hölling, Schlack, Petermann, Ravens-Sieberer & Mauz, 2014) ermittelt für den Zeitraum 2003 bis 2012 eine stabile Quote von rund 20 % aller Schüler_innen, die im Elternurteil Probleme im Verhalten aufweisen. Barkmann und Schulte-Markwort (2012) berechnen in ihrer Metaanalyse 33 epide-

miologischer Studien im Zeitraum von 1953 bis 2007 eine etwas geringere Prävalenz von 17.6 %. Anders als im statistischen Vergleich zwischen der KiGGS Basiserhebung und der KiGGS Welle 1, ermitteln Klipker, Baumgarten, Göbel, Lampert und Hölling (2018) mit 16.9 % einen signifikanten Rückgang der Gesamtprävalenz in KiGGS Welle 2. Die Autor_innen diskutieren den vermehrten Einsatz präventiver Maßnahmen als Erklärung für diesen positiven Trend.

Durch die Folgen der staatlich verfügten Maßnahmen zum Gesundheitsschutz im Zuge der COVID-19-Pandemie und die damit einhergehenden Einschränkungen, sowohl für die Gesamtbevölkerung als auch für Kinder und Jugendliche, sind die Effekte dieses Trends verschwunden. Die COPSY-Studie (Ravens-Sieberer et al., 2021) weist einen Anstieg von Verhaltensproblemen auf 30.4 % aus. Die befragten Eltern gaben an, dass sie sich unter anderem Unterstützung in Bezug auf das Verhalten ihrer Kinder und bei der Bewältigung schulischer Anforderungen wünschen. Ein Ansatz zur Verbesserung dieser Ausgangslage könnten schulische Fördermaßnahmen sein. Richard, Eichelberger, Döpfner und Hanisch (2015) verweisen in ihrer Übersichtsarbeit zu schulbasierten Interventionen bspw. auf das Good Behavior Game als wirksame Klassen-zentrierte Maßnahme zur Verbesserung des aufgabenbezogenen Verhaltens und zur Reduktion von störendem Verhalten.

Das Good Behavior Game

Zwei aktuelle Übersichtsarbeiten zu Schulinterventionen für Kinder mit Aufmerksamkeits- und aggressiven Verhaltensstörungen empfehlen übereinstimmend lerntheoretisch orientierte Maßnahmen (bspw. Tokensysteme; Fabiano & Pyle, 2019; Waschbusch, Breaux & Babinski, 2019), zu denen auch das Good Behavior Game (GBG; Barrish, Saunders & Wolf, 1969) zählt. In Deutschland ist die Maßnahme bekannt als KlasseKinderSpiel (Hillenbrand & Pütz, 2008). Konkret handelt es sich um eine präventive Intervention, die während des Unterrichts im Klassenraum durchgeführt wird und bei der es darum geht, sich an zuvor vereinbarte Verhaltensregeln zu halten. Zunächst wird die Klasse in zwei bis vier Teams aufgeteilt. Während der Spielzeit, die meist auf zehn bis zwanzig Minuten begrenzt ist, gelten einfach verständliche Regeln, die am besten mit den Kindern gemeinsam definiert werden (bspw. Ich sitze still, Ich arbeite leise, Ich melde mich). Jeder Verstoß wird als „Foul“ für die ganze Mannschaft mit einem Punkt geahndet. Das Team mit der geringsten Punktzahl am Ende der Spielzeit gewinnt eine Gruppenbelohnung. Demnach erhalten die Schüler_innen einer vorgegebenen Kleingruppe im Anschluss an die Spielphasen Gratifikationen, wenn sie gemeinsam zuvor definierte Verhaltensziele erreicht haben (weitere Informationen zur konkreten Umsetzung des GBG

sind im Methodenteil unter Durchführung enthalten). Im Kern kommen demnach kognitiv-behaviorale Strategien zum Einsatz, die zum Abbau störender Verhaltensweisen im Unterricht beitragen. Zugleich werden adäquate Verhaltensweisen positiv verstärkt. Der Teamcharakter unterstützt zudem die gegenseitige Unterstützung der Schüler_innen untereinander bei der Regeleinhaltung.

Die erste Wirksamkeitsevaluation zum GBG wurde bereits 1969 veröffentlicht (Barrish et al., 1969). Außerdem existieren zwei ältere Übersichtsarbeiten, in denen jedoch nicht ausschließlich schulische Settings berücksichtigt werden, die Effekte nicht quantifiziert werden und auch der Einfluss von Studienmerkmalen auf die Outcomes nicht untersucht wird (Embry, 2002; Tingstrom, Sterling-Turner & Wilczynski, 2006). Bis heute liegen zwei GBG-spezifische Metaanalysen vor, die diese Kriterien erfüllen (Bowman-Perrot, Burke, Zaini, Zhang & Vannest, 2015; Flower, McKenna, Bunuan, Muething & Vega, 2014). Während Flower et al. (2014) sowohl Gruppenuntersuchungen als auch Einzelfallstudien in ihrer Metaanalyse berücksichtigen, greifen Bowman-Perrott et al. (2015) ausschließlich auf Daten aus Einzelfallstudien zurück.

Metaanalytische Befunde

Flower et al. (2014) identifizieren insgesamt 22 in peer-reviewed Journals veröffentlichte Artikel, in denen Studienergebnisse zur Effektivität des GBG berichtet werden. Sechzehn der berücksichtigten Originalarbeiten beinhalten die Analyse von Daten aus Einzelfallstudien. Bei den übrigen sechs Untersuchungen handelt es sich um Kontrollgruppenstudien. Übergreifend zeigen die Befunde, dass der Einsatz des GBG störendes Schüler_innenverhalten im Unterricht reduziert sowie lernförderliches Verhalten (on-task) effektiv fördert. Faktoren, wie Alter der Schüler_innen, Schulung der Lehrkräfte oder Umsetzungstreue (fidelity), ergaben keine systematischen Unterschiede in Bezug auf die Wirksamkeit. Unter Berücksichtigung aller Gruppenuntersuchungen liegt der durchschnittliche Interventionseffekt auf den Rückgang herausfordernder Verhaltensweisen in schulischen Settings bei $d = 0.50$ (Flower et al., 2014). Eine auf 1439 Datenpunkten basierende mehrebenenanalytische Auswertung der Einzelfalldaten (hierarchical linear modeling) ergab zudem einen signifikanten und sofortigen Effekt des GBG auf herausforderndes Verhalten (-20.38% , $p < .01$), für den negativen Trend (treatment effect on trend) jedoch nicht (-0.03% , $p > .05$). Demnach nimmt das Problemverhalten der Schüler_innen zu Beginn der Durchführung des GBG unmittelbar ab und verbleibt im zeitlichen Verlauf auf einem niedrigeren Niveau.

Bowman-Perrot et al. (2015), die 21 Einzelfallstudien miteinander vergleichen, berichten übergreifend einen starken Interventionseffekt des GBG (Tau-U = .82; Tau-U nach Parker, Vannest, Davis & Sauber, 2011) mit einem

größeren Effekt auf den Rückgang von störendem Verhalten und Unaufmerksamkeit (off-task; $\text{Tau-U} = .81$) als auf on-task Verhalten ($\text{Tau-U} = .59$). Tau-U ist ein Maß für die Effektgröße, das auf der Überlappungsfreiheit zwischen den Messzeitpunkten der Ausgangslage (Baseline) und der Interventionsphase basiert. Eine Stärke ist, dass vorhandene Trends in der Baseline kontrolliert werden können (Bowman-Perrot et al., 2015). Der Index gibt den Prozentsatz der Messzeitpunkte als Dezimalzahl an, die sich im Laufe der Zeit verbessern (Parker et al., 2011).

Unterschiedliche Varianten des GBG sind vergleichbar effektiv. Zudem zeigen die Analysen, dass Schüler_innen mit Verhaltensstörungen oder einem entsprechenden Risiko am meisten profitieren ($\text{Tau-U} = .98$ vs. $\text{Tau-U} = .76$ bei Schüler_innen ohne Verhaltensprobleme).

Ergänzend zu den GBG-spezifischen metaanalytischen Befunden ordnen Chaffee, Briesch, Johnson und Volpe (2017) in ihrer Überblicksarbeit zu klassenweiten Interventionen in Regelschulsettings das GBG als hocheffiziente Maßnahme zur Unterstützung des Schüler_innenverhaltens ein. Die berichteten Befunde zur Häufigkeit von Verhaltensproblemen und zur Effektivität des GBG beziehen sich in erster Linie auf Schüler_innen in Regelklassen und ohne diagnostizierte psychiatrische und/oder Verhaltensstörungen.

Darüber hinaus belegen weitere Forschungsergebnisse, dass externalisierende Verhaltensprobleme (bspw. aggressives, oppositionelles und hyperaktives Verhalten) wirksam reduziert werden können (Kellam, Ling, Merisca, Brown & Ialongo, 1998; Kellam, Rebok, Ialongo & Mayer, 1994; Leflot, van Lier, Onghena & Colpin, 2010; Leflot, van Lier, Onghena & Colpin, 2013; Witvliet, van Lier, Cuijpers & Koot, 2009).

Einzelfallbefunde zur Wirksamkeit bei Schüler_innen mit externalisierenden Verhaltensproblemen

Die Befunde einer Einzelfallstudie von Groves und Austin (2017) belegen, dass sich sowohl Störverhalten als auch Unaufmerksamkeit bei vier Kindern mit psychiatrischen Störungen (im Alter von neun bis zehn Jahren) durch den Einsatz des GBG reduzieren. Joslyn und Vollmer (2020) berichten ebenfalls einen Rückgang des Störverhaltens für diese Zielgruppe, basierend auf Einzelfalldaten von drei Schulklassen (elementary-school und middle-school); wie auch Donaldson, Fisher und Kahng (2017), die Einzelfalldaten von zwölf Kindern (neun Vorschulkinder, drei Erstklässler_innen) mit stark ausgeprägtem Störverhalten auswerten. Eine weitere Einzelfallstudie belegt die Auswirkungen auf das on-task Verhalten von drei Kindern (jeweils acht Jahre alt) mit Aufmerksamkeitsproblemen (Pennington & McComas, 2017). Darüber hinaus weisen Groves und Austin (2019) den Rückgang negativer Interaktionen zwischen Schüler_innen mit diversen Formen

von psychiatrischen Störungen (beispielsweise Autismus-Spektrum-Störung, Aufmerksamkeits-Defizit-Hyperaktivitäts-Störung) während der Spielphasen nach. Die Stichprobe bestand aus insgesamt 13 Schüler_innen im Alter von neun bis 16 Jahren aus zwei Klassen. Die Einzelfalldaten wurden jeweils auf der Gruppenebene erhoben (unter anderem systematische Beobachtung positiver und negativer Peer-Interaktionen im Unterricht).

Bei der Sichtung der insgesamt zahlreich vorhandenen Einzelfallstudien fällt auf, dass überwiegend eine ganze Klasse oder Kleingruppe als Einzelfall betrachtet wird (bspw. Groves & Austin, 2019; Joslyn & Vollmer, 2020). Folglich lassen die Studienergebnisse in erster Linie Schlussfolgerungen zur Wirksamkeit des GBG auf der Klassenebene zu. Demnach liegen nur wenige gesicherte Informationen zu den Auswirkungen des GBG auf das Verhalten einzelner Schüler_innen vor. Insgesamt deuten die Befunde zur Wirksamkeit des GBG bei Schüler_innen mit externalisierenden Verhaltensproblemen darauf hin, dass die Maßnahme auch bei dieser Zielgruppe erfolgreich eingesetzt werden kann. Allerdings besteht keine Stichprobe der zitierten Studien aus Schüler_innen des deutschen Bildungssystems.

Untersuchungsfragen

Zentrales Anliegen der vorliegenden Untersuchung ist es, die Effektivität des GBG bei Grundschüler_innen mit externalisierenden Verhaltensproblemen zu überprüfen. (1) Es existieren kaum Effektivitätsnachweise, die die Wirksamkeit der Maßnahme bei Schüler_innen des deutschen Bildungssystems belegen. (2) International betrachtet ist der Kenntnisstand zur Effektivität der Maßnahme bei Schüler_innen mit externalisierenden Verhaltensproblemen auf individueller Ebene insgesamt gering. Daraus resultiert die Forschungsfrage: Trägt der regelmäßige Einsatz des GBG dazu bei, das Verhalten von Grundschüler_innen mit externalisierenden Verhaltensproblemen im Unterricht positiv zu beeinflussen?

Basierend auf den Befunden von Bowman-Perrot et al. (2015), Groves und Austin (2017) sowie Joslyn und Vollmer (2020) gehen wir davon aus, dass Schüler_innen mit externalisierenden Verhaltensproblemen, die fünfmal pro Woche am GBG teilnehmen, im Vergleich zur Ausgangslage bzw. Baseline (Verhalten des Kindes vor dem Beginn der Durchführung des GBG) sowohl signifikant seltener Störverhalten im Unterricht zeigen als auch ein deutlich verbessertes lernförderliches Verhalten während der Spielphasen. Wir vermuten zudem, dass sich der Einsatz des GBG auch auf das Verhalten der Kinder außerhalb der Spielphasen auswirkt. Bereits Leflot et al. (2010) berichten günstige Auswirkungen des GBG auf das Stör- und on-task

Verhalten von Primarstufenschüler_innen im Unterricht sowie auf weitere externalisierende Verhaltensprobleme. Die Befunde basieren auf Daten, die nicht ausschließlich während der Spielphasen erhoben wurden. Auch der von Flower et al. (2014) berichtete durchschnittliche Interventionseffekt auf herausfordernde Verhaltensweisen basiert auf analysierten Daten aus Kontrollgruppenstudien, in denen teilweise die allgemeine Verhaltensentwicklung der Proband_innen untersucht wurde.

Methoden

Stichprobe

Die kontrollierte Einzelfallstudie wurde in einer inklusiven Grundschule in einer Mittelstadt in Westdeutschland mit ca. 50 000 Einwohner_innen durchgeführt. Es handel-

te sich um eine vierzügige Primarstufenschule mit rund 300 Schüler_innen der Klassen eins bis vier. An der Studie nahmen alle vier dritten Klassen teil. Jede Klassenlehrkraft benannte fünf Schüler_innen mit den größten externalisierenden Verhaltensproblemen ($N = 20$). Sie erhielten vorab den Auftrag solche Kinder auszuwählen, die besonders häufig den Unterricht stören und Schwierigkeiten haben sich an die Regeln zu halten (verbale Unterrichtsstörungen, Verweigerung der Unterrichtsbeteiligung, verbale und körperliche Auseinandersetzungen) sowie / oder Konzentrations- und Hyperaktivitätsprobleme aufweisen (Unaufmerksamkeit, unruhiges Verhalten). Insgesamt besteht die Stichprobe aus einem Mädchen und 19 Jungen im Alter von acht ($n = 14$) bis neun Jahren ($n = 6$). Fünfzehn Kinder weisen einen Migrationshintergrund auf. Weitere demografische Informationen und Angaben zur Ausprägung von Verhaltensproblemen der Kinder sind in Tabelle 1 zusammengefasst. Die Daten zur Erhebung der Ausgangslage offenbaren große Unterschiede zwischen den Kindern in

Tabelle 1. Demografische Informationen zu den Schüler_innen und Ausprägung von Verhaltensproblemen (Erhebung der Ausgangslage)

Klasse	Schüler/in	Geschlecht	Alter in Jahren	Migrationshintergrund	Störverhalten M (SD)	Lernförderliches Verhalten M (SD)
3a	S1	m	8	ja	4.14 (0.38)	1.29 (0.95)
	S2	m	8	ja	4.86 (0.38)	1.00 (0.00)
	S3	m	8	ja	3.00 (0.58)	2.71 (0.49)
	S4	m	8	nein	2.43 (0.98)	2.86 (0.38)
	S5	m	8	nein	2.67 (0.52)	2.57 (0.53)
3b	S6	m	8	nein	1.67 (1.11)	3.27 (1.03)
	S7	m	9	ja	3.06 (0.77)	2.00 (0.97)
	S8	m	9	ja	1.56 (0.81)	2.12 (0.89)
	S9	m	9	nein	2.18 (1.13)	2.71 (0.85)
	S10	m	8	nein	2.41 (1.00)	2.47 (1.01)
3c	S11	m	9	ja	3.27 (0.80)	1.40 (0.63)
	S12	m	8	ja	3.13 (0.83)	1.53 (0.52)
	S13	m	8	ja	2.27 (0.88)	2.47 (0.64)
	S14	w	9	ja	3.33 (0.90)	2.00 (0.53)
	S15	m	8	ja	1.75 (0.50)	2.00 (0.82)
3d	S16	m	8	ja	3.50 (1.24)	1.25 (0.62)
	S17	m	8	ja	2.90 (0.91)	1.90 (0.83)
	S18	m	8	ja	1.81 (0.40)	1.50 (0.51)
	S19	m	8	ja	1.30 (0.57)	2.00 (0.73)
	S20	m	9	ja	2.19 (0.98)	2.52 (0.75)

Anmerkungen: M (SD): Mittelwerte und Standardabweichungen (Baseline Schultag). Skala von 0 = Verhalten trat nie auf bis 5 = Verhalten trat immer auf, Geschlecht: m = männlich, w = weiblich, Migrationshintergrund: mindestens ein Elternteil nicht in Deutschland geboren.

Bezug auf die Ausprägung externalisierender Verhaltensprobleme.

Alle Schüler_innen der vier Klassen nahmen am GBG teil. In jeder Klasse wurden ähnliche Strategien zur Verhaltensförderung und zur Klassenführung eingesetzt (bspw. Regeln, Unterrichtsorganisation und -methoden). Das GBG wurde täglich einmal von der Klassenlehrkraft zeitgleich zum Regelunterricht durchgeführt. In allen Klassen erfolgte die Durchführung in lehrkraftzentrierten Unterrichtssituationen, entweder in Einzelarbeits- oder Instruktionsphasen, da diese Situationen von den Lehrkräften in Bezug auf das Auftreten externalisierender Verhaltensprobleme als besonders problematisch angesehen wurden. Auch die aufgestellten Regeln waren in allen Klassen vergleichbar: 1. Ich sitze still, 2. Ich arbeite leise bzw. höre zu, 3. Ich melde mich.

Abhängige Variablen

Die Zielverhaltensweisen waren lernförderliches (LV) und störendes Verhalten (SV). Das lernförderliche Verhalten umfasste das aktive und engagierte Schüler_innenverhalten in Bezug auf die Mitarbeit im Unterricht (bspw. aktive Beteiligung am Unterricht, sich melden, konzentriert arbeiten, rechtzeitige Fertigstellung von Aufgaben). Störendes Verhalten wurde definiert als Verhalten, das den Unterricht stört oder das eigene Lernen oder das Lernen der anderen Kinder beeinträchtigt (bspw. in die Klasse rufen, rumalbern, unangemessene Seitengespräche, nicht auf dem Platz sitzen bleiben). Detaillierte Informationen zur Vorgehensweise bei der Erfassung dieser Merkmale sind im nächsten Abschnitt zum Erhebungsinstrument enthalten.

Erhebungsinstrument

Direct Behavior Rating (DBR)

Zur Messung von SV und LV wurden DBRs (Christ, Riley-Tillman & Chafouleas, 2009; Huber & Rietz, 2015) verwendet. Dabei handelt es sich um „eine Kombination aus einer direkten systematischen Verhaltensbeobachtung und einer Verhaltensbeurteilung mit Hilfe einer Ratingskala“ (Huber & Rietz, 2015, S.78). Casale, Hennemann und Grosche (2015) stellen fest, dass die Methode aufgrund ihrer Ökonomie und Flexibilität insbesondere für die Erfassung engmaschiger Verhaltensverläufe im Rahmen quantitativer Einzelfallstudien geeignet ist. Im Hinblick auf die Operationalisierung von SV und LV orientierten wir uns an Chafouleas (2011). Für beide abhängigen Variablen verwendeten wir eine Single-Item-Scale (SIS). Bei SIS wird ein breit definiertes Item genutzt, um ein allgemeines Verhaltensmerkmal zu repräsentieren (Christ et al., 2009).

Vorliegende Forschungsbefunde belegen die Generalisierbarkeit und Zuverlässigkeit über verschiedene Rater, Items und Beobachtungszeitpunkte hinweg (Casale, Grosche, Volpe & Hennemann, 2017; Casale, Hennemann, Volpe, Briesch & Grosche, 2015). Frühere Studien aus den USA weisen darüber hinaus die Zuverlässigkeit, Validität und Veränderungssensitivität von SIS in Bezug auf die fokussierten Verhaltensweisen (SV, LV) nach (bspw. Briesch, Chafouleas & Riley-Tillman, 2010; Chafouleas, Sanetti, Kilgus & Maggin, 2012). Analysen zur Messinvarianz belegen darüber hinaus die Reliabilität von DBR (Gebhardt, DeVries, Jungjohann, Casale, Gegenfurtner & Kuhn, 2019).

Der entwickelte Erhebungsbogen beinhaltet Definitionen der beiden Verhaltensmerkmale SV und LV in Verbindung mit einer Auflistung charakteristischer Beispiele, wie im Abschnitt zu den abhängigen Variablen beschrieben. Zur Beurteilung des Schüler_innenverhaltens mittels SIS durch die Lehrkräfte wird eine sechsstufige Likert-Skala verwendet. Die Ankreuzmöglichkeiten zu SV und LV sind: Das Verhalten trat nie (0), selten (1), manchmal (2), oft (3), sehr oft (4) und immer (5) auf. Die Instruktion lautet: Kreuzen Sie unmittelbar nach der Durchführung des GBG (Spielphase) jeweils die Zahl an, die am besten das Ausmaß des Verhaltens des Kindes beschreibt. Zusätzlich erfolgte täglich eine weitere Verhaltensbeurteilung durch die Klassenlehrkräfte direkt im Anschluss an die letzte Unterrichtsstunde zur Bewertung des gesamten Schultags.

Durchführung

Die Durchführung des GBG erfolgte an fünf Tagen pro Woche im Klassenverband während des Unterrichts. Auch wenn das GBG als einfach zu handhaben gilt, ist eine vorherige Schulung der Lehrkräfte und eine Betreuung während der Implementierung erforderlich (Poduska & Kurki, 2014). Hierzu erfolgte zunächst eine vierstündige Fortbildung mit dem gesamten Lehrer_innenkollegium. Inhaltlich wurden theoretische Grundlagen zur Prävention von Verhaltensauffälligkeiten thematisiert, das GBG detailliert vorgestellt (inkl. Filmbeispiel und Wirkmechanismen) sowie eine Arbeitsphase zur Vorbereitung der Umsetzung des GBG in der eigenen Klasse durchgeführt. Außerdem erhielten die Klassenlehrkräfte der dritten Klassen eine Schulung zum Einsatz von DBRs. Bei diesem Treffen wurden zunächst die beiden Verhaltensmerkmale SV und LV definiert und zur Konkretisierung mit beispielhaften Verhaltensweisen verknüpft (siehe Abschnitt zu den abhängigen Variablen). Außerdem wurde besprochen, dass die Beurteilung des Verhaltens immer durch die gleiche Person erfolgen muss und möglichst intuitiv. Anders als bei systematischen Verhaltensbeobachtungen, bei denen z. B. die Auftretenshäufigkeit bestimmter Verhaltensweisen gezählt wird, geht es bei DBRs darum, im Anschluss an eine zuvor festgelegte Unterrichtssituation, die Ausprägung be-

stimmter Verhaltensweisen in dieser Situation zu beurteilen (Casale, Hennemann, Huber & Grosche, 2015). Abschließend wurden der Zeitplan sowie die vorgesehenen Messzeitpunkte thematisiert und die Lehrkräfte erhielten eine Mappe, in der alle erforderlichen Erhebungsbögen enthalten waren.

Die Klassenlehrkräfte der dritten Klassen füllten die Erhebungsbögen während der Baseline täglich für alle fünf ausgewählten Zielkinder aus. Einmal direkt im Anschluss an die zehnminütige Unterrichtsphase in der das GBG zukünftig eingesetzt werden sollte und einmal am Ende des Schultags. Aufgrund schulorganisatorischer und krankheitsbedingter Umstände konnten die geplanten Messungen teilweise nicht realisiert werden (siehe Angaben zu fehlenden Messzeitpunkten in den Elektronischen Supplementen ESM1 und ESM2). Gründe hierfür waren z. B. die krankheitsbedingte Abwesenheit der Klassenlehrkraft oder des Kindes sowie Tage, an denen kein regulärer Unterricht stattfand (bspw. Ausflug, Sportfest, ganztägige schulinterne Fortbildung o.Ä.). Ab dem vorher festgelegten Interventionsstartpunkt führten die Klassenlehrkräfte das GBG durch. Die Lehrkräfte teilten ihre Klasse in zwei bis vier Teams ein, wobei sie die bestehenden Tischgruppen nutz-

ten. Anschließend überlegten sich die Mannschaften einen Namen für das eigene Team. Ein Plakat an der Tafel im vorderen Teil des Klassenzimmers listete die Regeln auf, die zuvor gemeinsam festgelegt wurden (z. B. Ich sitze still, Ich arbeite leise, Ich melde mich). Außerdem wurde für alle Teams gut sichtbar ein Timer in der Nähe der Tafel platziert und die Namen der Teams wurden an die Tafel geschrieben, um die Regelverstöße („Fouls“) während der Spielzeit zu markieren. In allen Klassenräumen wurde die Mannschaft mit den wenigsten „Fouls“ belohnt, oder mehrere Teams bei weniger als fünf Regelverstößen pro Mannschaft. Aufgrund unterschiedlicher Gegebenheiten in den verschiedenen Klassen wichen Regeln und Belohnungen minimal voneinander ab (z. B. Belohnung mit kleinen Sachgegenständen oder Aktivitäten). Diese Flexibilität war erforderlich, um die Umsetzung des GBG auf die spezifischen Bedingungen der jeweiligen Klasse anzupassen. Zudem deuten die Befunde von Flower et al. (2014) sowie Bowman-Perrot et al. (2015) darauf hin, dass entsprechende Modifikationen die Wirksamkeit nicht maßgeblich beeinflussen. Nach zehnminütiger Spielzeit in der zuvor festgelegten Unterrichtssituation zählten die Lehrkräfte die „Fouls“, benannten das Siegerteam und übergaben die Belohnungen. Während des

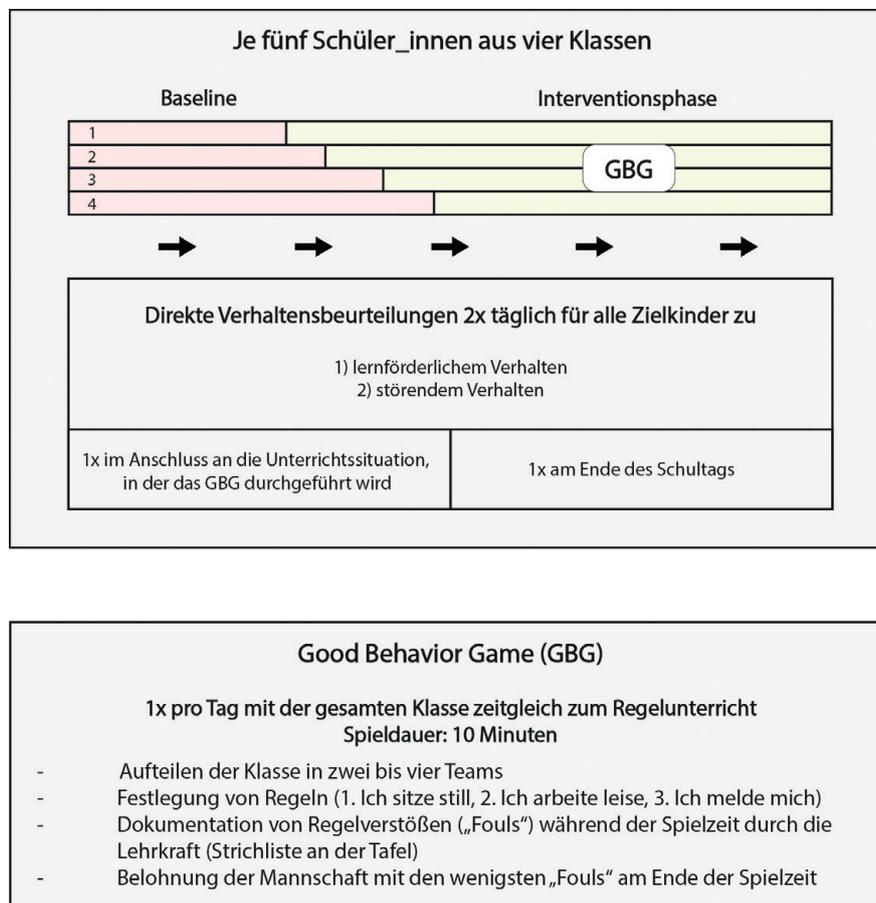


Abbildung 1. Forschungsdesign und Intervention.

kompletten Interventionszeitraums wurden die Erhebungsbögen wie oben beschrieben zweimal täglich für alle Zielkinder kontinuierlich ausgefüllt.

Um die Einhaltung der oben beschriebenen Vorgaben während der Durchführung sicherzustellen, wurden diese vorab ausführlich mit den Lehrkräften besprochen. Zusätzlich erhielt jede Lehrkraft mit der Mappe, in der auch die Erhebungsbögen enthalten waren, eine Checkliste mit den wichtigsten Kriterien (Spieldauer und Häufigkeit, Anzeige von „Fouls“, Belohnung der Gewinner). Zusätzlich gab es regelmäßige Austauschtreffen mit einem Projektmitarbeiter zur Klärung von Fragen und Problemen während der Durchführung.

Forschungsdesign und Hypothesen

Die Hypothesenprüfung erfolgte mittels eines Multiplen-Grundraten-Versuchsplans mit A-B Design (Jain & Spieß, 2012). Jede Klasse wurde nach dem Zufallsprinzip einem vorab festgelegten Interventionsstartpunkt zugewiesen (Kratochwill & Levin, 2010). In der Baseline (A-Phase) wurde die Ausprägung der abhängigen Variablen für die ausgewählten Kinder pro Klasse über zehn bis 25 Messzeitpunkte hinweg erhoben, ohne dass eine spezifische Intervention erfolgte. Die Datenpunkte der A-Phase dienen als Grundrate mit den Zielen der Deskription, Prognose und des Vergleichs (Jain & Spieß, 2012). In der Interventionsphase (B-Phase) startete dann die Umsetzung des GBG und die Zielvariablen wurden weiterhin über 30 bis 42 Erhebungszeitpunkte hinweg konstant erfasst (siehe Angaben zu Messzeitpunkten in ESM1 und ESM2).

Pro Schultag waren zwei Messzeitpunkte vorgesehen. Einer direkt im Anschluss an die Unterrichtssituation, die für den Einsatz des GBG ausgewählt wurde, und ein zweiter direkt im Anschluss an die letzte Schulstunde zur Bewertung des gesamten Schultags. Dieses Vorgehen ermöglicht (1) Aussagen über die Verhaltensentwicklung der Kinder in den Spielphasen sowie (2) die Kontrolle von Auswirkungen des GBG auf das Verhalten der Schüler_innen außerhalb der Spielphasen (Transfereffekte).

Das zentrale Anliegen unseres Vorhabens ist die Beantwortung der Frage, ob das GBG das Verhalten der ausgewählten Kinder im Unterricht positiv beeinflusst. Konkret gehen wir davon aus, dass der tägliche Einsatz des GBG einen Anstieg des LV und einen Rückgang des SV bei den ausgewählten Schüler_innen mit externalisierenden Verhaltensproblemen während der Spielphasen bewirkt (Hypothese 1). Zudem vermuten wir, dass sich das Verhalten der Zielkinder auch außerhalb der Spielphasen verbessert (Hypothese 2).

Das gewählte Forschungsdesign stellt eine Möglichkeit dar, um diese Hypothesen zu prüfen (Jain & Spieß, 2012).

In Abgrenzung zu Kontrollgruppenuntersuchungen werden individuelle Verhaltensverläufe engmaschig erfasst sowie das Verhalten eines Kindes vor Beginn der Intervention als Referenz herangezogen (individuelle Bezugsnormorientierung). Durch die Nutzung eines Multiplen-Grundraten-Versuchsplans über Personen können zudem diverse Störfaktoren minimiert werden (wie zwischenzeitliches Geschehen oder Reifung).

Datenanalyse

Um die Auswirkungen des GBG auf SV und LV auf Einzellebene und fallübergreifend zu untersuchen, wurden verschiedene Verfahren zur Datenanalyse genutzt. Die Abbildungen 2 und 3 geben einen Überblick zu den Effekten der Förderung für jedes Zielkind. Zur Beurteilung des Fördererfolgs auf Einzellebene wurde das Effektstärkemaß *Nonoverlap of all pairs* (NAP; Alresheed, Hott & Bano, 2013; Parker & Vannest, 2009) verwendet. Dieser Index vergleicht paarweise alle Datenpunkte aus Baseline und Interventionsphase. Beim Ergebnis handelt es sich um den Anteil der positiven und neutralen Trends aller möglichen Vergleichspaare, wobei neutrale Trends nur halb so stark gewichtet werden. Eine detaillierte Beschreibung des NAP inklusive Anwendungsbeispiel ist enthalten in Spilles und Hagen (2019). Nach Parker und Vannest (2009) sind mittlere Effekte durch Werte von 66 % bis 92 % und starke Effekte durch Werte ab 93 % gekennzeichnet. Werte ab 50 % repräsentieren eine tendenzielle Verbesserung. Die in ESM1 und ESM2 enthaltene deskriptive Statistik für jeden Einzelfall beinhaltet zudem Angaben zur Anzahl der Messzeitpunkte sowie Mittelwerte und Standardabweichungen für Baseline und Interventionsphase (GBG-Phase) pro Kind (Manolov & Moeyaert, 2017).

Aufgrund der in Phasen unterteilten Zeitreihendaten (Baseline A und Intervention B) nutzen wir zur fallübergreifenden Schätzung der Effekte einen stückweisen Regressionsansatz (piecewise regression; Huitema & Mckean, 2000). Dieses Vorgehen ermöglicht die Kontrolle von Entwicklungstrends in den Daten (Trendeffekte) sowie die Unterscheidung zwischen kontinuierlichen (Slope-Effekte) und unmittelbaren (Level-Effekte) Interventionseffekten. Beim Trendeffekt handelt es sich um einen Entwicklungseffekt, der als kontinuierliche Veränderung über alle Messungen hinweg interpretiert werden kann. Der Level-Effekt repräsentiert einen Interventionseffekt, der zu Beginn der Interventionsphase zu einer sofortigen und konstanten Veränderung führt. Der Slope-Effekt ist ein zweiter Interventionseffekt, der sich kontinuierlich mit dem Beginn der Intervention entwickelt (Wilbert, 2021). Wilbert, Börnert-Ringleb und Lüke (2022) konstatieren, dass es sich hierbei um einen Ansatz handelt, der ein hohes Potenzial aufweist,

um die Effekte von Interventionen in Einzelfallstudien adäquat zu identifizieren.

Die genestete Datenstruktur erforderte außerdem eine Mehrebenenenerweiterung (hierarchical piecewise regression; Moeyaert, Ferron, Beretvas & Van den Noortgate, 2014; Rindskopf & Ferron, 2014; Van den Noortgate & Onghena, 2003) mit Messzeitpunkten auf Ebene 1 in Subjekten auf Ebene 2. Die Mehrebenenanalysen wurden als Random Intercept- und Random Slope-Modell angelegt, mit allen drei Parametern (trend, slope und level effects) als fixe und zufällige Faktoren. Zur Signifikanztestung der zufälligen Effekte wurde für jeden Random Slope-Faktor ein Likelihood-Ratio-Test durchgeführt, bei dem das vollständige Modell mit einem Modell ohne diesen Faktor verglichen wurde. Alle Analysen wurden mit R (R Core Team, 2021) und dem Paket scan (Wilbert & Lüke, 2021) durchgeführt.

Ergebnisse

Abbildung 2 gibt einen Überblick zu den Effekten der Förderung auf Einzelfallebene. Angegeben ist das Effektstärkemaß NAP für die Verhaltensentwicklung in Bezug auf SV und LV während der Good Behavior Game-Phase (GBG-Phase). Die Effekte für die beiden abhängigen Variablen

variieren zwischen 32.33% und 97.45% für SV und zwischen 31.82% und 90.32% für LV. Lediglich bei vier Kindern gibt es keine tendenzielle Verbesserung in Bezug auf SV. Mindestens mittlere Effekte liegen bei neun Schüler_innen vor. Im Hinblick auf LV profitieren 13 Kinder mit mindestens mittleren Effekten. Demgegenüber zeigt der NAP ebenfalls bei vier Schüler_innen keine positive Tendenz an.

Aus Abbildung 3 sind analog zur Auflistung der Effekte für SV und LV während der GBG-Phase, die Befunde zur allgemeinen Verhaltensentwicklung (Schultag) zu entnehmen. Die Ermittlung des NAP für alle Einzelfälle ergibt ebenfalls eine große Spannweite, die zwischen 17.17% und 91.11% für SV und zwischen 17.30% und 95.76% für LV liegt. Im Hinblick auf SV gibt es bei 17 Kindern mindestens mittlere Effekte. Nur bei drei Schüler_innen kann keine tendenzielle Verbesserung festgestellt werden. In Bezug auf LV profitieren 14 Kinder mit mindestens mittleren Effekten. Demgegenüber zeigt der NAP nur bei einem Kind keine positive Tendenz an.

Mehrebenenanalysen

Zur Ermittlung der Gesamteffekte des GBG unter Berücksichtigung aller 20 Einzelfälle verwendeten wir Mehrebenenanalysen mit Messzeitpunkten auf Ebene 1 in Subjek-

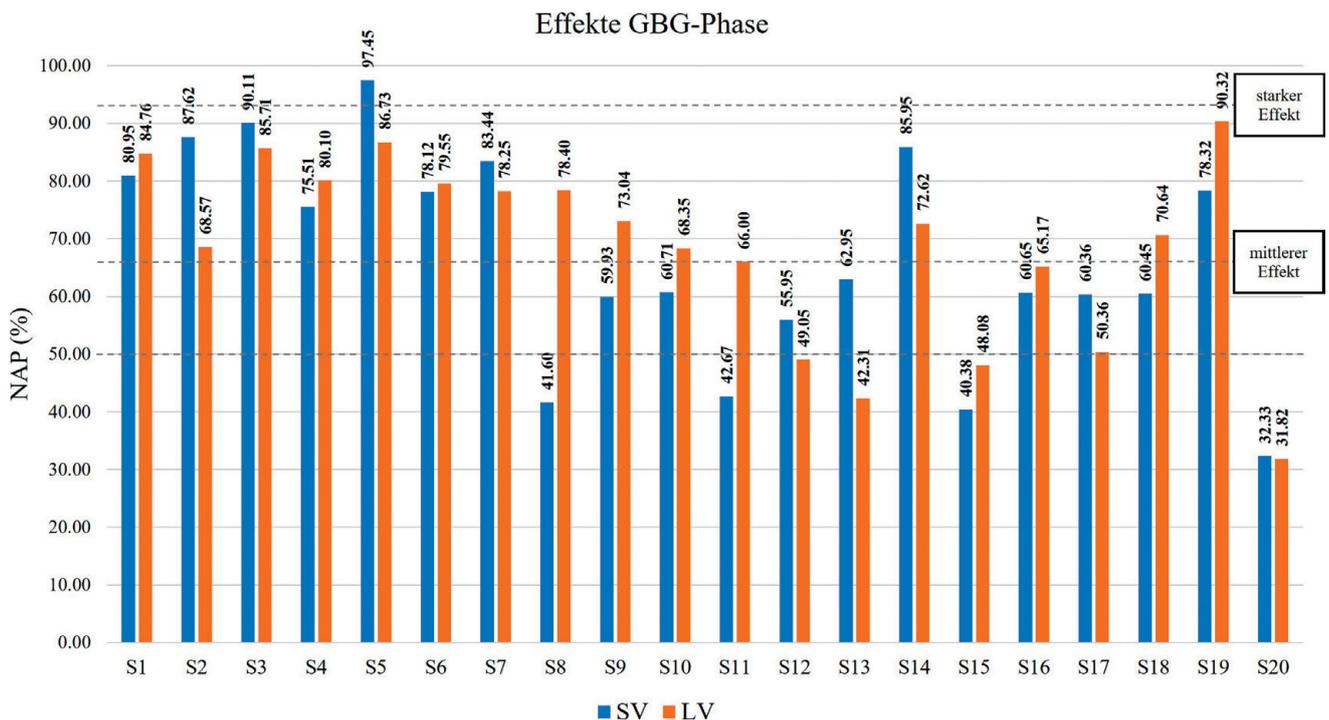


Abbildung 2. Effekte Good Behavior Game-Phase (GBG-Phase). Anmerkungen: S1-S20 = Einzelfälle, NAP = Nonoverlap of all pairs (Parker & Vannest, 2009), SV = Störverhalten, LV = lernförderliches Verhalten, Werte im Bereich von 50 – 100% = tendenzielle Verbesserung.

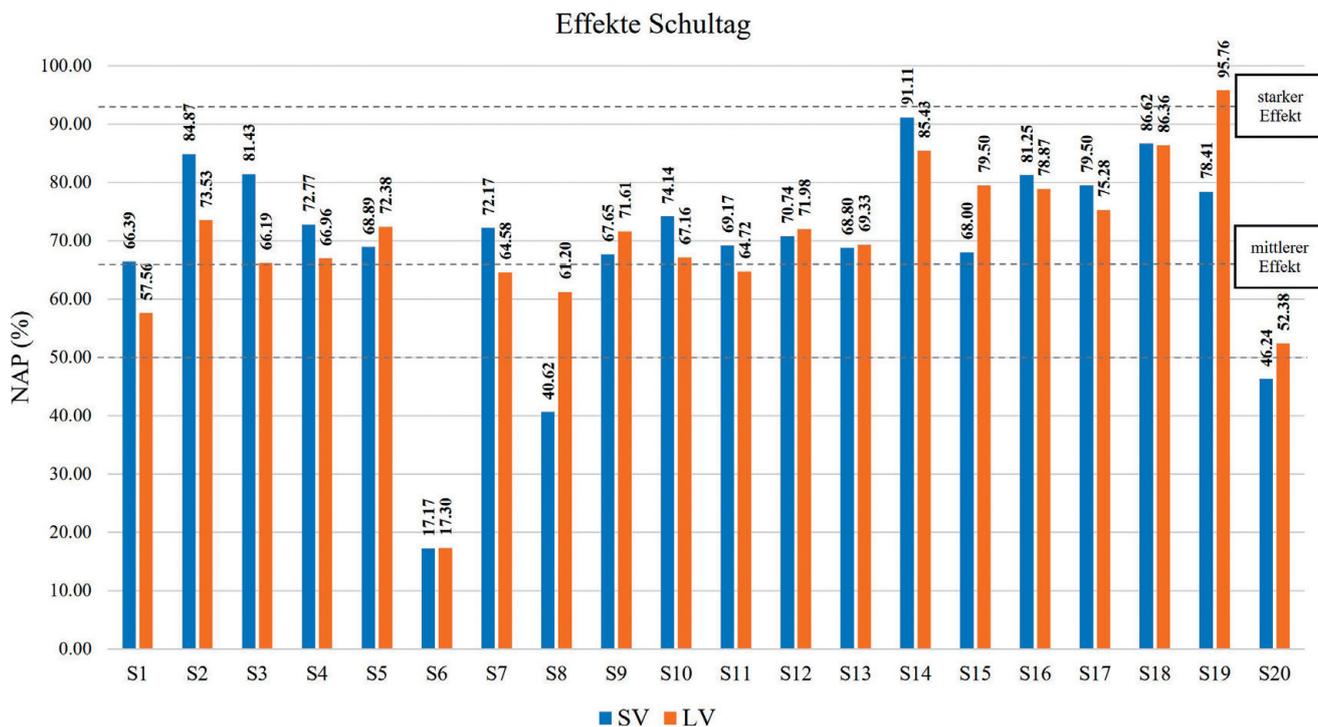


Abbildung 3. Effekte in Bezug auf den Schultag. Anmerkungen: S1-S20 = Einzelfälle, NAP = Nonoverlap of all pairs (Parker & Vannest, 2009), SV = Störverhalten, LV = lernförderliches Verhalten, Werte im Bereich von 50 – 100 % = tendenzielle Verbesserung.

ten auf Ebene 2. Die resultierenden Modellkoeffizienten sind in Tabelle 2a (GBG-Phase) und Tabelle 2b (Schultag) dargestellt.

Die Analysen ergeben jeweils signifikante Level-Effekte für SV und LV. Bei diesem Effekt handelt es sich um einen Interventionseffekt, der zu Beginn der Interventionsphase zu einer sofortigen und konstanten Verhaltensänderung führt (Wilbert, 2021). Im Durchschnitt sank das SV um 0.404 Punkte in der GBG-Phase ($p = .020$) und um 0.401 Punkte in Bezug auf den gesamten Schultag ($p = .028$). Das LV stieg um 0.432 Punkte in der GBG-Phase ($p = .006$) und um 0.442 Punkte ($p = .004$) außerhalb der Spielphasen (Schultag). Die Slope-Effekte für SV und LV erreichen jeweils keine statistische Signifikanz. Hierbei handelt es sich um einen zweiten Interventionseffekt, der sich kontinuierlich mit dem Beginn der Intervention entwickelt (Steigungseffekt).

Diskussion

Das zentrale Anliegen der vorliegenden Untersuchung bestand darin, die Effektivität des GBG im Hinblick auf LV und SV für Grundschüler_innen mit externalisierenden Verhaltensauffälligkeiten zu bestätigen. In Ergänzung zu früheren Einzelfallstudien betrachteten wir dabei nicht

nur die individuelle Verhaltensentwicklung während der Spielphasen (GBG-Phase), sondern zusätzlich den gesamten Schultag, um zu überprüfen, ob der Einsatz des GBG das Verhalten der Schüler_innen auch außerhalb der Spielphasen positiv beeinflusst.

Die Datenanalysen auf individueller Ebene ergeben ein heterogenes Bild. Während der GBG-Phase variieren die ermittelten Effekte (NAP) zwischen ca. 30 % und über 90 % für die beiden abhängigen Variablen. Im Hinblick auf SV kann nach den Konventionen von Parker und Vannest (2009) für acht der 20 Fälle ein mittlerer Effekt und für einen Fall ein starker Effekt nachgewiesen werden. In Bezug auf LV liegen mittlere Effekte bei 14 Schüler_innen vor. Demnach ist der Anteil der Schüler_innen, die während der GBG-Phase im LV von der Intervention profitieren, etwas größer. Ferner fällt auf, dass insgesamt sechs der elf Kinder, bei denen keine Effekte im SV nachgewiesen wurden (NAP < 66 %), bereits zu Beginn der Untersuchung eher gering ausgeprägte Verhaltensprobleme zeigten (SV ≤ 1.50; Bodeneffekte).

Im Hinblick auf die Verhaltensentwicklung der Schüler_innen außerhalb der Spielphasen (Schultag) ist im Vergleich zur GBG-Phase bemerkenswert, dass der Anteil der Kinder größer ist, die im SV profitieren. Auf dieser Dimension kann für 17 Fälle ein mittlerer Effekt nachgewiesen werden. Im LV profitieren insgesamt 13 Schüler_innen mit einem mittleren Effekt und ein Kind mit einem starken

Tabelle 2a. Feste und zufällige Effekte für Stör- und Lernverhalten in der Good Behavior Game-Phase

Parameter	Feste Effekte				Zufällige Effekte		
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>Estimated SD</i>	<i>L</i>	<i>p</i>
Störverhalten (SV)							
Intercept	2.179	0.256	8.503	0.000	1.055	80.08	0.000
Trend	-0.009	0.011	-0.847	0.397	0.022	3.33	0.505
Level Phase B	-0.404	0.173	-2.335	0.020	0.510	7.09	0.131
Slope Phase B	-0.003	0.011	-0.252	0.801	0.021	1.05	0.902
Lernförderliches Verhalten (LV)							
Intercept	1.928	0.190	10.137	0.000	0.727	38.29	0.000
Trend	0.021	0.012	1.757	0.079	0.034	7.51	0.111
Level Phase B	0.432	0.158	2.734	0.006	0.424	5.74	0.219
Slope Phase B	-0.010	0.016	-0.609	0.543	0.052	12.34	0.015

Tabelle 2b. Feste und zufällige Effekte für Stör- und Lernverhalten in Bezug auf den Schultag

Parameter	Feste Effekte				Zufällige Effekte		
	<i>B</i>	<i>SE</i>	<i>t</i>	<i>p</i>	<i>Estimated SD</i>	<i>L</i>	<i>p</i>
Störverhalten (SV)							
Intercept	2.784	0.273	10.189	0.000	1.135	78.43	0.000
Trend	-0.024	0.017	-1.373	0.170	0.063	26.37	0.000
Level Phase B	-0.401	0.183	-2.196	0.028	0.592	12.37	0.015
Slope Phase B	0.018	0.020	0.903	0.367	0.074	26.19	0.000
Lernförderliches Verhalten (LV)							
Intercept	2.107	0.203	10.367	0.000	0.822	60.09	0.000
Trend	-0.003	0.013	-0.250	0.803	0.044	16.48	0.002
Level Phase B	0.442	0.153	2.884	0.004	0.485	13.06	0.011
Slope Phase B	0.013	0.015	0.854	0.393	0.052	15.56	0.004

Effekt von der Intervention. Demnach ist im Vergleich zur GBG-Phase der Anteil der Proband_innen für die keine Effekte nachgewiesen werden können ($NAP < 66\%$) insgesamt etwas geringer (SV: $n = 3$; LV: $n = 6$).

Die Befunde der Mehrebenenanalysen unter Berücksichtigung aller 20 Einzelfälle weisen sowohl für die GBG-Phase als auch für den gesamten Schultag signifikante Level-Effekte auf beiden Dimensionen (SV und LV) nach. Während sich das LV verbessert, geht das SV zurück. Die berechneten Slope-Effekte hingegen sind nicht signifikant. Demnach zeigt sich ein unmittelbarer Interventionseffekt auf das Verhalten der Schüler_innen, ohne dass sich ein statistisch bedeutsamer Trend im Interventionszeitraum

abzeichnet. Diese Feststellung gilt sowohl für die Verhaltensentwicklung während der Spielphasen (GBG-Phase) als auch für den gesamten Schultag. Während Bowman-Perrot et al. (2015) stärkere Effekte bei der Reduzierung von SV im Vergleich zur Verbesserung des LV berichten, weisen unsere Analysen eine vergleichbare Veränderung beider Verhaltensdimensionen nach. In Übereinstimmung mit Flower et al. (2014) zeigt sich in unseren Daten ebenfalls kein statistisch bedeutsamer Rückgang von SV im zeitlichen Verlauf der Interventionsphase.

Eine wesentliche Erweiterung bisheriger Forschungen zum GBG stellen die Einzelfallbefunde zur Verhaltensentwicklung der Schüler_innen außerhalb der Spielphasen dar.

Demnach ist von Transfereffekten des GBG im Hinblick auf den Rückgang von SV und die Verbesserung des LV auszugehen. Die ermittelten Effekte sind vergleichbar mit der Verhaltensänderung in den Spielphasen (GBG-Phase).

Insgesamt weisen die Befunde auf Chancen und Grenzen des GBG hin. Bei einem Großteil der Schüler_innen zeigen sich sowohl im SV als auch im LV positive Effekte. Jeweils 70 % der Kinder der Stichprobe konnten ihr individuelles LV verbessern. Das SV reduzierte sich sogar bei 85 % der Schüler_innen außerhalb der Spielphasen (Schultag), während der GBG-Phase allerdings lediglich bei 45 %. Dies könnte mit der Auswahl der Spielsituationen zusammenhängen, die für manche Kinder eine geringere Herausforderung darstellten. Vermutlich fiel es einigen Kindern von vornherein leichter, sich in diesen Unterrichtssituationen an die Regeln zu halten, wodurch sie folglich weniger Unterstützung durch das GBG erhielten als das bei anderen Schüler_innen der Fall war.

Die Befunde bestätigen und erweitern bisherige Forschungsergebnisse, die darauf hindeuten, dass das GBG geeignet ist, um Verhaltensprobleme von Schüler_innen mit Verhaltensauffälligkeiten im Unterricht zu reduzieren (Bowman-Perrot et al., 2015; Donaldson et al., 2017; Groves & Austin, 2017; Pennington & McComas, 2017). Überdies zeigen die Ergebnisse in Übereinstimmung mit Ashworth, Humphrey, Lendrum und Hennessey (2020), dass nicht alle Schüler_innen vom GBG profitieren. Die Autor_innen gehen davon aus, dass die Maßnahme nicht intensiv genug ist, um eine Wirkung auf die Hochrisikogruppe zu erzielen. Demnach könnte es sinnvoll sein, das GBG insbesondere bei Kindern, die multiplen Entwicklungsrisiken ausgesetzt sind, in eine umfangreichere Interventionsstrategie einzubetten, bei der verschiedene Maßnahmen sinnvoll miteinander kombiniert werden (bspw. Hanisch et al., 2019; Waschbusch et al., 2019).

Limitationen

Aufgrund der großen Variabilität zwischen den Einzelfällen (deutliche Unterschiede in Bezug auf die Ausgangswerte der Kinder zur Ausprägung von Verhaltensproblemen; siehe Tab. 1), sollte die Auswahl der Schüler_innen zukünftig nicht ausschließlich den Lehrkräften überlassen werden, obwohl Studienbefunde darauf hinweisen, dass Lehrkräfte kompetent darin sind, Schüler_innen mit Verhaltensproblemen in ihrer Klasse zu identifizieren (Lane & Menzies, 2005). Zwar ist es grundsätzlich sinnvoll, die subjektive Einschätzung von Lehrkräften bei der Auswahl von Kindern ihrer Klasse mit besonders herausfordernden Verhaltensweisen einzubeziehen, zur zusätzlichen Absicherung sollten jedoch ergänzend etablierte Screenings (bspw. SDQ; Goodman, 1997) eingesetzt werden. Eine weitere

Möglichkeit stellt die Auswahl von Kindern mit bereits vorliegenden Störungsdiagnosen dar. Darüber hinaus fällt die extrem schiefe Geschlechterverteilung der Stichprobe auf (1 Mädchen vs. 19 Jungen), weshalb die Resultate in erster Linie für männliche Probanden aussagekräftig sind.

Relativiert werden müssen die Befunde außerdem aufgrund der Tatsache, dass die DBRs sowohl für die GBG-Phase als auch für den gesamten Schultag von den gleichen Klassenlehrkräften durchgeführt wurden. Es könnte sein, dass die Beurteilung des Verhaltens der Kinder in den Spielphasen die Beurteilung des gesamten Schultags beeinflusst hat. Demnach sollten die DBRs in Folgeuntersuchungen von unterschiedlichen Personen durchgeführt werden, um die gefundenen Transfereffekte abzusichern.

Aus methodischer Sicht resultiert aus hohen Werten im SV bzw. geringen Werten im LV bei einigen Schüler_innen ein minimaler Spielraum für Verbesserungen (Boden- bzw. Deckeneffekte; Fabiano & Pyle, 2019; Ho & Yu, 2015). Zudem wirkt sich die Nutzung einer SIS zur Messung des Störverhaltens limitierend auf die Aussagekraft unserer Studie aus. Während bereits nach fünf Messungen zuverlässige Ergebnisse zum Lernverhalten vorliegen (Volpe & Briesch, 2012), sind für aussagekräftige Daten zum Störverhalten mehr als 20 Messungen erforderlich (Casale et al., 2017; Volpe & Briesch, 2012). Demnach sollte in Folgeuntersuchungen eine MIS zur Beurteilung von SV verwendet werden, auch wenn damit ein etwas höherer Aufwand bei der Datenerhebung verbunden ist.

Die teilweise signifikanten Effektivitätsunterschiede zwischen den Fällen, insbesondere in Bezug auf den gesamten Schultag, legen die Vermutung nahe, dass die Wirkung des GBG durch weitere Merkmale beeinflusst wird, die im genutzten Modell nicht berücksichtigt sind. Sowohl Bowman-Perrot et al. (2015) als auch Ashworth et al. (2020) berichten bspw. unterschiedliche Interventionseffekte in Abhängigkeit von der Ausprägung von Verhaltensproblemen. Dies spricht für die Berücksichtigung weiterer Level 2-Faktoren in Folgeuntersuchungen, um diese Unterschiede zu verstehen. Allerdings ist hierfür eine Mindestanzahl von 30 Einzelfällen erforderlich (Wilbert, 2021).

Relevanz für die Praxis

Die Analyse individueller Entwicklungsverläufe hilft dabei Schüler_innen zu identifizieren, die zusätzliche Unterstützung benötigen (Donaldson et al., 2017). Zahlreiche Untersuchungen betrachten jedoch eine Klasse bzw. Gruppe als Einzelfall (bspw. Groves & Austin, 2019; Joslyn & Vollmer, 2020). Dadurch sind kaum Rückschlüsse auf die individuelle Verhaltensentwicklung einzelner Schüler_innen mit Verhaltensproblemen möglich. Das von uns gewählte Vorgehen zeigt einen Weg auf, wie dieses Problem gelöst wer-

den kann. Es ermöglicht, die Auswirkungen des GBG auf das Verhalten einzelner Schüler_innen sowie die Wirksamkeit über alle Fälle hinweg zu untersuchen.

Insgesamt deuten die Ergebnisse unserer Studie darauf hin, dass das GBG die Verhaltensentwicklung von Grundschüler_innen mit Verhaltensproblemen im deutschen Bildungssystem fördert. Sie können von dieser einfach zu handhabenden Intervention profitieren. Die Befunde sollten Lehrkräfte dazu ermutigen, das GBG in ihrer Klasse einzusetzen. Diese Schlussfolgerung wird durch die Befunde zur Verhaltensentwicklung der Schüler_innen außerhalb der Spielphasen zusätzlich gestützt. Aus pädagogischer Perspektive handelt es sich bei den gefundenen Transfer-effekten um Interventionswirkungen mit großer praktischer Bedeutsamkeit, denn die Maßnahme soll Kindern dabei helfen ihr Verhalten im Unterricht insgesamt zu verbessern, und dies nicht nur während der Spielphasen.

Elektronische Supplemente (ESM)

Die elektronischen Supplemente sind mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1024/2235-0977/a000410>.

ESM1. Deskriptive Statistik der Einzelfalldaten zur Good Behavior Game-Phase.

ESM2. Deskriptive Statistik der Einzelfalldaten zum Schultag.

Literatur

- Alresheed, F., Hott, B. L. & Bano, C. (2013). Single Subject Research: A Synthesis of Analytic Methods. *Journal of Special Education Apprenticeship*, 2, 1–18.
- Ashworth, E., Humphrey, N., Lendrum, A. & Hennessey, A. (2020): Beyond "what works": A mixed-methods study of intervention effect modifiers in the Good Behavior Game. *Psychology in the Schools*, 57, 222–246. <https://doi.org/10.1002/pits.22312>
- Barkmann, C. & Schulte-Markwort, M. (2012). Prevalence of emotional and behavioural disorders in German children and adolescents: a meta-analysis. *Journal of Epidemiology and Community Health* (1979-), 66, 194–203. <https://doi.org/10.1136/jech.2009.102467>
- Barrish, H., Saunders, M. & Wolf, M. (1969). Good behavior game. Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, 2, 119–124.
- Bowman-Perrott, L., Burke, M., Zaini, S., Zhang, N. & Vannest, K. (2015). Promoting Positive Behavior Using the Good Behavior Game: A Meta-Analysis of Single-Case Research. *Journal of Positive Behavior Interventions*, 18, 1–11. <https://doi.org/10.1177/1098300715592355>
- Briesch, A. M., Chafouleas, S. M. & Riley-Tillman, T. C. (2010). Generalizability and Dependability of Behavior Assessment Methods to Estimate Academic Engagement: A Comparison of Systematic Direct Observation and Direct Behavior Rating. *School Psychology Review*, 39, 408–421. <https://doi.org/10.1080/02796015.2010.12087761>
- Casale, G., Grosche, M., Volpe, R. J. & Hennemann, T. (2017). Zuverlässigkeit von Verhaltensverlaufsdiagnostik über Rater und Messzeitpunkte bei Schülern mit externalisierenden Verhaltensproblemen. *Empirische Sonderpädagogik*, 9, 143–164.
- Casale, G., Hennemann, T. & Grosche, M. (2015). Zum Beitrag der Verlaufsdiagnostik für eine evidenzbasierte sonderpädagogische Praxis am Beispiel des Förderschwerpunktes der emotionalen und sozialen Entwicklung. *Zeitschrift für Heilpädagogik*, 7, 325–334.
- Casale, G., Hennemann, T., Huber, C. & Grosche, M. (2015) Testgütekriterien der Verlaufsdiagnostik von Schülerverhalten im Förderschwerpunkt Emotionale und soziale Entwicklung. *Heilpädagogische Forschung*, 41, 37–54.
- Casale, G., Hennemann, T., Volpe, R. J., Briesch, A. M. & Grosche, M. (2015). Generalisierbarkeit und Zuverlässigkeit von Direkten Verhaltensbeurteilungen des Lern- und Arbeitsverhaltens in einer inklusiven Grundschulklasse. *Empirische Sonderpädagogik*, 7, 258–268.
- Chaffee, R. K., Briesch, A. M., Johnson, A. H. & Volpe, R. J. (2017). A Meta-Analysis of Class-Wide Interventions for Supporting Student Behavior. *School Psychology Review*, 46, 149–164. <https://doi.org/10.17105/SPR-2017-0015.V46-2>
- Chafouleas, S. M. (2011). Direct behavior rating: A review of the issues and research in its development. *Education And Treatment of Children*, 34, 575–591. <https://doi.org/10.1353/etc.2011.0034>
- Chafouleas, S. M., Sanetti, L. M., Kilgus, S. P. & Maggin, D. M. (2012). Evaluating sensitivity to behavioral change using direct behavior rating single-item scales. *Exceptional Children*, 78, 491–505. <https://doi.org/10.1177/001440291207800406>
- Christ, T. J., Riley-Tillman, T. C. & Chafouleas, S. M. (2009). Foundation for the development and use of direct behavior rating (DBR) to assess and evaluate student behavior. *Assessment for Effective Intervention*, 34, 201–213. <https://doi.org/10.1177/1534508409340390>
- Donaldson, J. M., Fisher, A. B. & Kahng, S. W. (2017). Effects of the Good Behavior Game on individual student behavior. *Behavior Analysis: Research and Practice*, 17, 207–216. <https://doi.org/10.1037/bar0000016>
- Embry, D. D. (2002). The Good Behavior Game: A Best Practice Candidate as a Universal Behavioral Vaccine. *Clinical Child and Family Psychology Review*, 5, 273–297. <https://doi.org/10.1023/A:1020977107086>
- Fabiano, G. & Pyle, K. (2019). Best practices in school mental health for attention-deficit/hyperactivity disorder: A framework for intervention. *School Mental Health*, 11, 72–91. <https://doi.org/10.1007/s12310-018-9267-2>
- Flower, A., McKenna, J. W., Bunuan, R. L., Muething, C. S. & Vega, R. (2014). Effects of the Good Behavior Game on Challenging Behaviors in School Settings. *Review of Educational Research*, 84, 546–571. <https://doi.org/10.3102/0034654314536781>
- Gebhardt, M., DeVries, J., Jungjohann, J., Casale, G., Gegenfurtner, A. & Kuhn, J.-T. (2019). Measurement invariance of a direct behavior rating multi item scale across occasions. *Social Sciences*, 8, 46. <https://doi.org/10.3390/socsci8020046>
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>
- Groves, E. A. & Austin, J. L. (2017). An evaluation of interdependent and independent group contingencies during the Good Behavior Game. *Journal of Applied Behavior Analysis*, 50, 552–566. <https://doi.org/10.1002/jaba.393>

- Groves, E. A. & Austin, J. L. (2019): Does the Good Behavior Game evoke negative peer pressure? Analyses in primary and secondary classrooms. *Journal of Applied Behavior Analysis*, 52, 3 – 16. <https://doi.org/10.1002/jaba.513>
- Hanisch, C., Casale, G., Volpe, R. J., Briesch, A. M., Richard, S., Meyer, H. et al. (2019). Gestufte Förderung in der Grundschule: Konzeption eines mehrstufigen, multimodalen Förderkonzeptes bei expansivem Problemverhalten. *Prävention und Gesundheitsförderung*, 14, 237 – 241. <https://doi.org/10.1007/s11553-018-0700-z>
- Hillenbrand, C. & Pütz, K. (2008). *KlasseKinderSpiel. Spielerisch Verhaltensregeln lernen*. Hamburg: Edition Körber Stiftung.
- Ho, A. D. & Yu, C. C. (2015). Descriptive statistics for modern test score distributions: Skewness, kurtosis, discreteness, and ceiling effects. *Educational and Psychological Measurement*, 75, 365 – 388. <https://doi.org/10.1177/0013164414548576>.
- Hölling, H., Schlack, R., Petermann, F., Ravens-Sieberer, U. & Mauz, E. (2014). Psychische Auffälligkeiten und psychosoziale Beeinträchtigungen bei Kindern und Jugendlichen im Alter von 3 bis 17 Jahren in Deutschland – Prävalenz und zeitliche Trends zu 2 Erhebungszeitpunkten (2003 – 2006 und 2009 – 2012). *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 57, 807 – 819. <https://doi.org/10.1007/s00103-014-1979-3>
- Huber, C. & Rietz, C. (2015). Direct Behavior Rating (DBR) als Methode zur Verhaltensverlaufdiagnostik in der Schule: Ein systematisches Review von Methodenstudien. *Empirische Sonderpädagogik*, 2, 75 – 98.
- Huitema, B. E. & Mckean, J. W. (2000). Design Specification Issues in Time-Series Intervention Models. *Educational and Psychological Measurement*, 60, 38 – 58. <https://doi.org/10.1177/00131640021970358>
- Jain, A. & Spieß, R. (2012). Versuchspläne der experimentellen Einzelfallforschung. *Empirische Sonderpädagogik*, 4, 211 – 245.
- Joslyn, P. R. & Vollmer, T. R. (2020): Efficacy of teacher-implemented good behavior game despite low treatment integrity. *Journal of Applied Behavior Analysis*, 53, 465 – 474. <https://doi.org/10.1002/jaba.614>
- Kellam, S. G., Ling, X., Merisca, R., Brown, C. H. & Ialongo, N. (1998). The effect of the level of aggression in the first grade classroom on the course and malleability of aggressive behavior into middle school. *Development and Psychopathology*, 10, 165 – 185. <https://doi.org/10.1017/S0954579498001564>
- Kellam, S. G., Rebok, G. W., Ialongo, N. & Mayer, L. S. (1994). The course and malleability of aggressive behavior from early first grade into middle school: Results of a developmental epidemiologically-based preventive trial. *Journal of Child Psychology and Psychiatry*, 35, 259 – 281. <https://doi.org/10.1111/j.1469-7610.1994.tb01161.x>
- Klipker, K., Baumgarten, F., Göbel, K., Lampert, T. & Hölling, H. (2018). Psychische Auffälligkeiten bei Kindern und Jugendlichen in Deutschland – Querschnittergebnisse aus KiGGS Welle 2 und Trends. *Journal of Health Monitoring*, 37 – 45. <https://doi.org/10.17886/RKI-GBE-2018-077>
- Kratochwill, T. R. & Levin, J. R. (2010). Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue. *Psychological Methods*, 15, 124 – 144. <https://doi.org/10.1037/a0017736>
- Lane, K. L. & Menzies, H. M. (2005). Teacher-identified students with and without academic and behavioral concerns: Characteristics and responsiveness. *Behavioral Disorders*, 31, 65 – 83.
- Leflot, G., van Lier, P. A. C., Onghena, P. & Colpin, H. (2010). The role of teacher behavior management in the development of disruptive behaviors: An intervention study with the Good Behavior Game. *Journal of Abnormal Child Psychology*, 38, 869 – 882.
- Leflot, G., van Lier, P. A. C., Onghena, P. & Colpin, H. (2013). The role of children's on-task behavior in the prevention of aggressive behavior development and peer rejection: A randomized controlled study of the Good Behavior Game in Belgian elementary classrooms. *Journal of School Psychology*, 51, 187 – 199. <https://doi.org/10.1016/j.jsp.2012.12.006>
- Manolov, R. & Moeyaert, M. (2017). Recommendations for Choosing Single-Case Data Analytical Techniques. *Behavior Therapy*, 48, 97 – 114. <https://doi.org/10.1016/j.beth.2016.04.008>
- Moeyaert, M., Ferron, J. M., Beretvas, S. N. & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, 52, 191 – 211. <https://doi.org/10.1016/j.jsp.2013.11.003>
- Parker, R. I. & Vannest, K. (2009). An Improved Effect Size for Single-Case Research: Nonoverlap of All Pairs. *Behavior Therapy*, 40, 357 – 367. <https://doi.org/10.1016/j.beth.2008.10.006>
- Parker, R. I., Vannest, K. J., Davis, J. L. & Sauber, S. B. (2011). Combining non-overlap and trend for single-case research: Tau-U. *Behavior Therapy*, 42, 284 – 299. <https://doi.org/10.1016/j.beth.2010.08.006>
- Pennington, B. & McComas, J. J. (2017). Effects of the Good Behavior Game across classroom contexts. *Journal of Applied Behavior Analysis*, 50, 176 – 180. <https://doi.org/10.1002/jaba.357>
- Poduska, J. M. & Kurki, A. (2014). Guided by Theory, Informed by Practice: Training and Support for the Good Behavior Game, a Classroom-Based Behavior Management Strategy. *Journal of Emotional and Behavioral Disorders*, 22, 83 – 94. <https://doi.org/10.1177/1063426614522692>
- R Core Team (2021). *R: A language and environment for statistical computing*. Zugriff am 14.02.2022. <https://www.r-project.org>
- Ravens-Sieberer, U., Kaman, A., Otto, C., Adedeji, A., Napp, A.-K., Becker, M. et al. (2021). Seelische Gesundheit und psychische Belastungen von Kindern und Jugendlichen in der ersten Welle der COVID-19-Pandemie – Ergebnisse der COPSY-Studie. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 64, 1512 – 1521. <https://doi.org/10.1007/s00103-021-03291-3>
- Richard, S., Eichelberger, I., Döpfner, M. & Hanisch, C. (2015). Schulbasierte Interventionen bei ADHS und Aufmerksamkeitsproblemen: Ein Überblick. *Zeitschrift für Pädagogische Psychologie*, 29, 5 – 18. <https://doi.org/10.1024/1010-0652/a000141>
- Rindskopf, D. & Ferron, J. M. (2014). Using multilevel models to analyze single-case design data. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (S. 221 – 246). <https://doi.org/10.1037/14376-008>
- Spilles, M. & Hagen, T. (2019). Schulpraktische Erhebungs- und Auswertungsmöglichkeiten von Einzelfalldaten. *Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF)*, Nr. 3, 1 – 14.
- Tingstrom, D. H., Sterling-Turner, H. E. & Wilczynski, S. M. (2006). The Good Behavior Game: 1969 – 2002. *Behavior Modification*, 30, 225 – 253. <https://doi.org/10.1177/0145445503261165>
- Van den Noortgate, W. & Onghena, P. (2003). Hierarchical linear models for the quantitative integration of effect sizes in single-case research. *Behavior Research Methods, Instruments & Computers*, 35, 1 – 10. <https://doi.org/10.3758/BF03195492>
- Volpe, R. J. & Briesch, A. M. (2012). Generalizability and Dependability of Single-Item and Multiple-Item Direct Behavior Rating Scales for Engagement and Disruptive Behavior. *School Psychology Review*, 41, 246 – 261. <https://doi.org/10.1080/02796015.2012.12087506>
- Waschbusch, D. A., Breaux, R. P. & Babinski, D. E. (2019). School-based interventions for aggression and defiance in youth: a framework for evidence-based practice. *School Mental Health*, 11, 92 – 105. <https://doi.org/10.1007/s12310-018-9269-0>
- Wilbert, J. & Lüke, T. (2021). *Single-Case Data Analyses for Single and Multiple AB Designs (scan) (Version 0.53) [R]*. Zugriff am 06.10.2021. <https://r-forge.r-project.org/projects/scan>

- Wilbert, J. (2021). *Analyzing Single-Case Data with R and Scan*. Zugriff am 14.02.2022. <https://jazznbass.github.io/scan-Book/index.html>
- Wilbert, J., Börnert-Ringleb, M. & Lüke, T. (2022). Statistical Power of Piecewise Regression Analyses of Single-Case Experimental Studies Addressing Behavior Problems. *Frontiers in Education*, 7, Article 917944. <https://doi.org/10.3389/feduc.2022.917944>
- Witvliet, M., van Lier, P. A. C., Cuijpers, P. & Koot, H. M. (2009). Testing links between childhood positive peer relations and externalizing outcomes through a randomized controlled intervention study. *Journal of Consulting and Clinical Psychology*, 77, 905–915. <https://doi.org/10.1037/a0014597>



Dr. Tobias Hagen

Lehrstuhl für Erziehungshilfe
und sozial-emotionale
Entwicklungsförderung
Department Heilpädagogik
und Rehabilitation
Humanwissenschaftliche Fakultät
Universität zu Köln
Klosterstraße 79c
50931 Köln
Deutschland
t.hagen@uni-koeln.de

Historie

Manuskript eingereicht: 22.02.2022

Manuskript angenommen: 24.02.2023

Onlineveröffentlichung: 09.03.2023

Förderung

Open-Access-Veröffentlichung ermöglicht durch die Universität zu Köln.

ORCID

Tobias Hagen

 <https://orcid.org/0000-0002-5465-1568>

Jannik Nitz

 <https://orcid.org/0000-0003-0358-8973>

Dennis Christian Hövel

 <https://orcid.org/0000-0003-0933-2661>

Thomas Hennemann

 <https://orcid.org/0000-0003-4961-8680>