

Potsdamer Zentrum für empirische Inklusionsforschung (ZEIF), 2017, Nr. 1

## „Nützt es was oder nützt es nichts?“ -

### Direct Behavior Rating (DBR) als diagnostische Methode zur zeitnahen Überprüfung des Fördererfolgs bei unterrichtlichem Schülerinnen- und Schülerverhalten



*Gino Casale  
Universität zu Köln*

**Zusammenfassung:** Die Umsetzung einer evidenzbasierten pädagogischen Praxis setzt diagnostische Verfahren voraus, die zur Überprüfung des Fördererfolgs im Einzelfall geeignet sind. In diesem Kontext haben sich Methoden der Verlaufsdiagnostik als geeignet erwiesen. Während für den Bereich des akademischen Lernens bereits einige solcher Verfahren entwickelt und evaluiert wurden, fehlen im Bereich des Unterrichtsverhaltens von Schülerinnen und Schülern noch vergleichbare Verfahren, die für eine engmaschige Erfassung des Verhaltens geeignet sind. Das Direct Behavior Rating (DBR) hat sich international und in ersten Studien auch in Deutschland als potenzielles Verfahren für diese Zwecke bewährt. Der vorliegende Beitrag stellt die Methode konzeptionell vor. Zentrale Aspekte des aktuellen Forschungsstandes zur Testgüte werden – vor dem Hintergrund der praktischen Relevanz – präsentiert. Die Methode wird anhand eines Fallbeispiels exemplarisch für die praktische Umsetzung angewendet.

**Schlagwörter:** Verhaltensdiagnostik, Evidenzbasierte Praxis, Direct Behavior Rating

**Abstract:** Assessment tools which evaluate a particular student's response to intervention present a key element of evidence-based practice in schools. Progress-monitoring measures have been advocated as appropriate assessment tools for these purposes. However, there is still a lack of those tools in the behavioral domain. Direct Behavior Rating (DBR) has been proposed as a promising tool to timely and frequently assess a student's school-based behavior. This article presents the conceptual and theoretical framework of DBR. Key aspect of the current state of evidence is presented, focusing on practical issues. DBR is exemplified by a hypothetical usage in a case example.

**Keywords:** Behavior Assessment, Evidence-based Practice, Direct Behavior Rating

## Einleitung

Viele Kinder und Jugendliche zeigen in der Schule Verhaltensweisen, die unterrichtliche Prozesse und Abläufe beeinträchtigen (Forness, Kim, & Walker, 2012). Epidemiologische Studien über die Auftretenshäufigkeit beziffern den Anteil an Kindern und Jugendlichen mit problematischem Verhalten auf ca. 12 – 20% (Barkmann & Schulte-Markwort, 2012; Hölling et al., 2014). Diese Verhaltensweisen gefährden den individuellen und kollektiven Lern- und Entwicklungserfolg und stellen eine enorme Belastung für die Mitschülerinnen und Mitschüler sowie für die Lehrkräfte im Klassenraum dar, insbesondere wenn sie externalisierend, d. h. nach außen gerichtet sind (Haller et al., 2016; Makarova, Herzog, & Schönbacher, 2014). Zusätzlich kann problematisches Verhalten in der Schule mit erheblichen Kurz- und Langzeitfolgen, wie z. B. der Entwicklung einer manifesten psychischen Störung, akademischen Scheitern oder sozialer Ablehnung, einhergehen (Costello, Copeland, & Angold, 2011; Huber, 2009; Krull, Wilbert, & Hennemann, 2014; Reinke, Herman, Petras, & Ialongo, 2008).

Diese Erkenntnisse verdeutlichen den hohen Bedarf an schulischen Verhaltensfördermaßnahmen, die frühzeitig eingesetzt werden und bestenfalls positiv auf das Verhalten der Schülerinnen und Schüler (SuS) wirken. In diesem Kontext wird häufig auf evidenzbasierte Maßnahmen, deren Wirksamkeit wissenschaftlich nachgewiesen wurde, verwiesen, um

die Wahrscheinlichkeit des Lern- und Entwicklungserfolges zu erhöhen (Hillenbrand, 2015). Der wissenschaftliche Nachweis über die Wirksamkeit von Fördermaßnahmen kann z. B. über groß angelegte Gruppenuntersuchungen erbracht werden, d. h. die Förderung wird an einer Gruppe mit sehr vielen Personen (z. B. SuS einer Grundschule) durchgeführt. Vor und nach der Förderung wird die Ausprägung bestimmter Merkmale (z. B. das Sozialverhalten) erfasst und mit der Ausprägung einer anderen Gruppe, die nicht gefördert wird, verglichen. Eine Maßnahme wird dann als wirksam bezeichnet, wenn sich die SuS der geförderten Gruppe besser entwickelt haben als die SuS der Gruppe ohne Förderung und wenn die Verbesserung mit hoher Wahrscheinlichkeit auf die Umsetzung der Maßnahme und nicht auf den Zufall zurückgeführt werden kann.

Wenngleich die Befunde aus Gruppenuntersuchungen für die Beurteilung der Wirksamkeit einer Fördermaßnahme enorm wichtig und ein wesentliches Merkmal einer evidenzbasierten Praxis in der Schule sind, gibt es dennoch keine Garantie dafür, dass solche Fördermaßnahmen auch im Einzelfall erfolgreich sind (Grosche, 2014). Daher ist es wichtig, den Fördererfolg auch im Einzelfall zu prüfen und systematisch Daten zu erheben, die zeitnah Aufschluss darüber geben, ob sich ein/e Schüler/in durch eine Intervention positiv entwickelt oder nicht. Im optimalen Fall ergänzen sich beide Ansätze gegenseitig, d. h. durch in Gruppenuntersuchungen positiv evaluierte

Fördermaßnahmen wird auch im Einzelfall ein Fördererfolg erreicht. Dieses Konzept wird auch als „Evidenzbasierung im Einzelfall“ (Casale, Hennemann, & Grosche, 2015, S. 327) bezeichnet. In der praktischen Umsetzung wird darunter verstanden, dass engmaschig Daten hinsichtlich eines spezifischen Verhaltensmerkmals erhoben werden, die dazu genutzt werden, pädagogisch-praktische Entscheidungen zu treffen (z. B. die Entscheidung eine Förderung beizubehalten oder sie zu modifizieren).

## **Instrumente der Verhaltensverlaufsdagnostik**

Damit zeitnah und engmaschig die notwendigen Informationen, die für die Beurteilung über den Erfolg einer Fördermaßnahme nötig sind, erhoben werden können, bedarf es entsprechender diagnostischer Verfahren, die für diese Zwecke geeignet sind (Casale, Hennemann, Huber, & Grosche, 2015). Zu diesem Zweck hat sich die engmaschige Verlaufsdagnostik als vielversprechende diagnostische Vorgehensweise erwiesen (Wilbert, 2014). In der Verlaufsdagnostik werden Schülerinnen und Schüler in relativ kurzen Abständen mit gleichen oder parallelen Aufgaben getestet, um Entwicklungsverläufe der Kinder über die Zeit (anhand der individuellen Bezugsnorm) abbilden zu können. Wird in der Verlaufsdagnostik eine positive Entwicklung nachgewiesen, wird die Förderung fortgeführt. Lässt sich jedoch keine positive Entwicklung während der Förderung nachweisen, wird die Förderung besser auf die Lern- und

Entwicklungsbedürfnisse des Kindes ausgerichtet (Grosche, 2014; Huber & Grosche, 2012).

Während für den Bereich des akademischen Lernens bereits einige Verfahren der Verlaufsdagnostik existieren und auch im deutschsprachigen Raum wissenschaftlich überprüft wurden (z.B. Diehl, 2012; Strathmann & Klauer, 2012; Walter, 2014), steckt die Entwicklung und Evaluation von Methoden der Verhaltensverlaufsdagnostik noch in den Kinderschuhen (Huber & Rietz, 2015b). Traditionelle und überprüfte Verfahren, wie z. B. standardisierte Verhaltensbeobachtungen oder umfangreiche Entwicklungstests sind zu aufwändig für den engmaschigen Nutzung im Rahmen der Verlaufsdagnostik (Casale et al., 2015). Daher wurde die in den USA entwickelte Methode des Direct Behavior Ratings (kurz: DBR; zu Deutsch: Direkte Verhaltensbeurteilung; DVB) in den deutschen Sprachraum übertragen und in ersten Studien wissenschaftlich überprüft (Casale, Hennemann, Volpe, Briesch, & Grosche, 2015; Huber & Rietz, 2015a). Im Folgenden sollen die Methode sowie der Forschungsstand hinsichtlich der Testgüte vorgestellt werden.

## **Direct Behavior Rating als Methode der Verhaltensverlaufsdagnostik**

Die praktische Umsetzung von DBR (ein Beispiel findet sich in Abbildung 1) sieht so aus, dass ein interessierendes Verhaltensmerkmal (z. B. Melden, wenn es Fragen gibt) festgelegt wird, welches durch eine Förderung (z. B. Verstärker-

plan) in einer bestimmten Situation im Schulalltag (z. B. individuelle Stillarbeitsphase) verbessert werden soll. Direkt im Anschluss an die jeweilige Situation wird das entsprechende Verhalten auf einer Ratingskala von der Lehrkraft beurteilt. Die Beurteilung kann zu jeder weiteren Situation – auch mehrmals täglich – wiederholt werden, so dass bereits nach wenigen Beurteilungen bestimmte Entwicklungstrends im Verhalten der Schülerin oder des Schülers dokumentiert werden können. Aus diesen Trends, die sich anschaulich mit Liniendiagrammen darstellen lassen, können Rückschlüsse auf die Wirksamkeit der Förderung (in diesem Beispiel: des Verstärkerplans) gezogen werden (Grosche, 2014). Verbessert sich etwa das Verhalten während der Förderung, wird die Förderung beibehalten. Ist keine positive Verhaltensänderung feststellbar, wird die Förderung entsprechend angepasst. Die Methode liefert somit einen Beitrag zur engmaschigen und ressourcenorientierten Lern- und Förderbegleitung der Kinder und Jugendlichen und zur Evidenzbasierung von konkreten Fördermaßnahmen im Einzelfall.

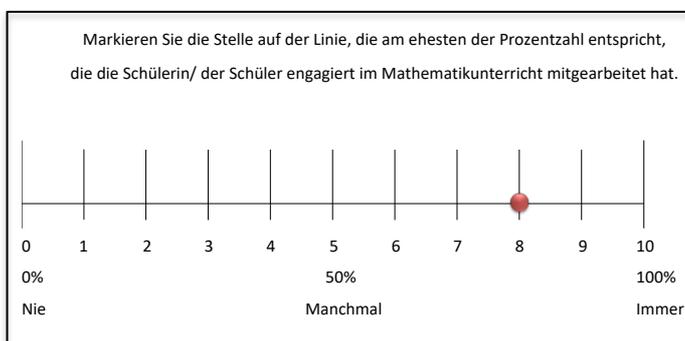


Abbildung 1. Beispiel eines Direct Behavior Ratings zur Beurteilung des lernförderlichen Verhaltens im Mathematikunterricht (entnommen und modifiziert von: [www.directbehaviorrating.com](http://www.directbehaviorrating.com))

Sowohl in ihrer formalen Gestaltung als auch in ihrer konkreten Anwendung sind DBR äußerst flexibel und praktikabel, da sie für verschiedenste Zwecke, Kontexte und Verhaltensweisen eingesetzt werden können (Christ, Riley-Tillman, & Chafouleas, 2009; Abbildung 2). So kann prinzipiell jedes beobachtbare Verhalten (z. B. störendes Verhalten, Lern- und Arbeitsverhalten), das gefördert werden soll oder aber ein Problem darstellt, beurteilt werden. Hierbei kann eine globale Verhaltensdimension (z. B. „Störendes Verhalten“) oder eine bzw. mehrere spezifische Verhaltensdimension(en) (z. B. „Benutzt Schimpfwörter“, „Verliert die Geduld“ oder „Lästert über andere Kinder“) in den Blick genommen werden. Bei der Verwendung eines einzigen Items spricht man von Single-Item-Skalen (SIS). Die Direkte Verhaltensbeurteilung mit mehreren Items wird dementsprechend als Multi-Item-Skala (MIS) bezeichnet. Sowohl die Lehrkraft als auch andere pädagogische Fachkräfte können das Verhalten beobachten und anschließend einschätzen. Wichtig dabei ist nur, dass die Beurteilungen konsistent durch dieselbe Person erfolgen. Dies schließt nicht aus, dass auch mehrere Personen parallel das gleiche Kind beurteilen, insbesondere weil erste deutschsprachige Forschungsbefunde darauf hinweisen, dass die Genauigkeit der Urteile höher ist, wenn mehrere Personen das Verhalten einer Schülerin oder eines Schülers parallel beurteilen (Casale et al., 2015). Die Situationen, in denen das Verhalten eingeschätzt wird, können nach dem individuellen Ziel der Förderung variieren (Christ et al., 2009). So kann das

Schüler/innenverhalten in Unterrichtsphasen (z. B. Gruppenarbeit oder individuelle Stillarbeit), aber auch in Pausen- und Übergangszeiten (z. B. Gang in die Sporthalle) fokussiert werden. Es muss lediglich beobachtbar und in der jeweiligen Situation von Relevanz sein (Christ et al., 2009).

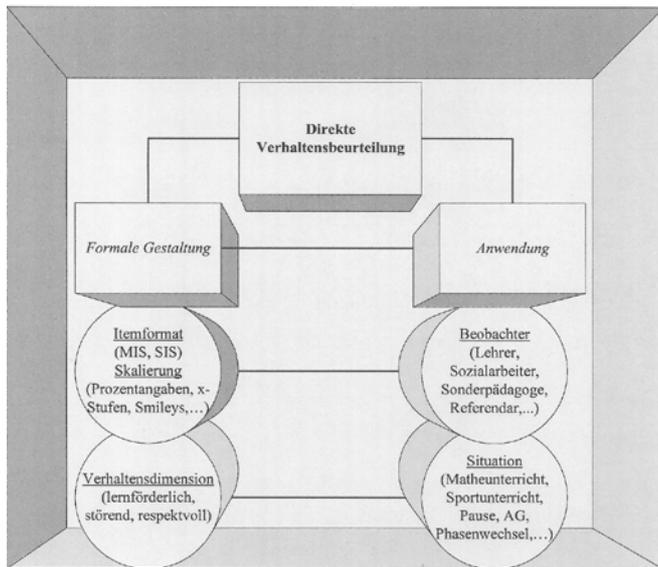


Abbildung 2. Potentiale zur Anwendung von DBR (entnommen aus Casale, Hennemann & Grosche, 2015, S. 330)

### Anwendungsbeispiel

Lisa ist Schülerin in einer vierten Klasse einer Grundschule. Ihrem Klassenlehrer fällt schon seit längerer Zeit auf, dass Lisa in Gruppenarbeitsphasen ständig stört, d. h. sie unterbricht ihre Mitschülerinnen und Mitschüler, ist nur wenig kompromissbereit, schreit häufig herum und fragt nur unangemessen nach Hilfe. Der Klassenlehrer hat auf einer Fortbildung von der Methode der Verhaltensverlaufdiagnostik mit Direct Behavior Rating erfahren und überlegt sich, die Methode einzusetzen, um das Verhalten von Lisa systematisch zu erfassen. So kann er

seinen Anfangsverdacht („Lisa stört ständig in Gruppenarbeitsphasen“) strukturieren und systematisieren. Er entwickelt eine DBR-Multi-Item-Skala mit vier Items („Unterbricht andere“, „Geht keine Kompromisse ein“, „Schreit herum“ und „Fragt unangemessen nach Hilfe“) und einer sechsstufigen Skala (0 = Verhalten tritt nie auf bis 5 = Verhalten tritt immer auf). Über den Zeitraum von einer Woche schätzt er zweimal täglich, nach jeder Gruppenarbeitsphase, diese vier Verhaltensweisen ein. Um eine möglichst umfassende Einschätzung zu bekommen, bittet er die Sonderschullehrerin in seinem Co-Team sowie den in der Klasse eingesetzten Praktikanten ebenfalls, die Ratings auszufüllen. Alle drei Beurteiler/innen kommen zu der Einschätzung, dass Lisa in Gruppenarbeitsphasen stark stört, wengleich sich die absoluten Werte alle unterscheiden, d. h. alle drei Personen andere Maßstäbe zur Verhaltensbeurteilung anlegen (Abbildung 3).

Nach einer Woche wendet der Klassenlehrer das Klasse-Kinder-Spiel (KKS) an, eine lerntheoretisch fundierte und evidenzbasierte Maßnahme, die über spielerische Elemente prosoziales Verhalten auf Gruppenebene fördert (siehe Hillenbrand & Pütz, 2008 für weitere Informationen). In den Gruppenarbeitsphasen erfassen sowohl der Klassenlehrer als auch die Sonderschullehrerin und der Praktikant das störende Verhalten weiterhin. Nach bereits einer Woche lässt sich visuell eine starke Verbesserung des Verhaltens feststellen (Abbildung 3), was den

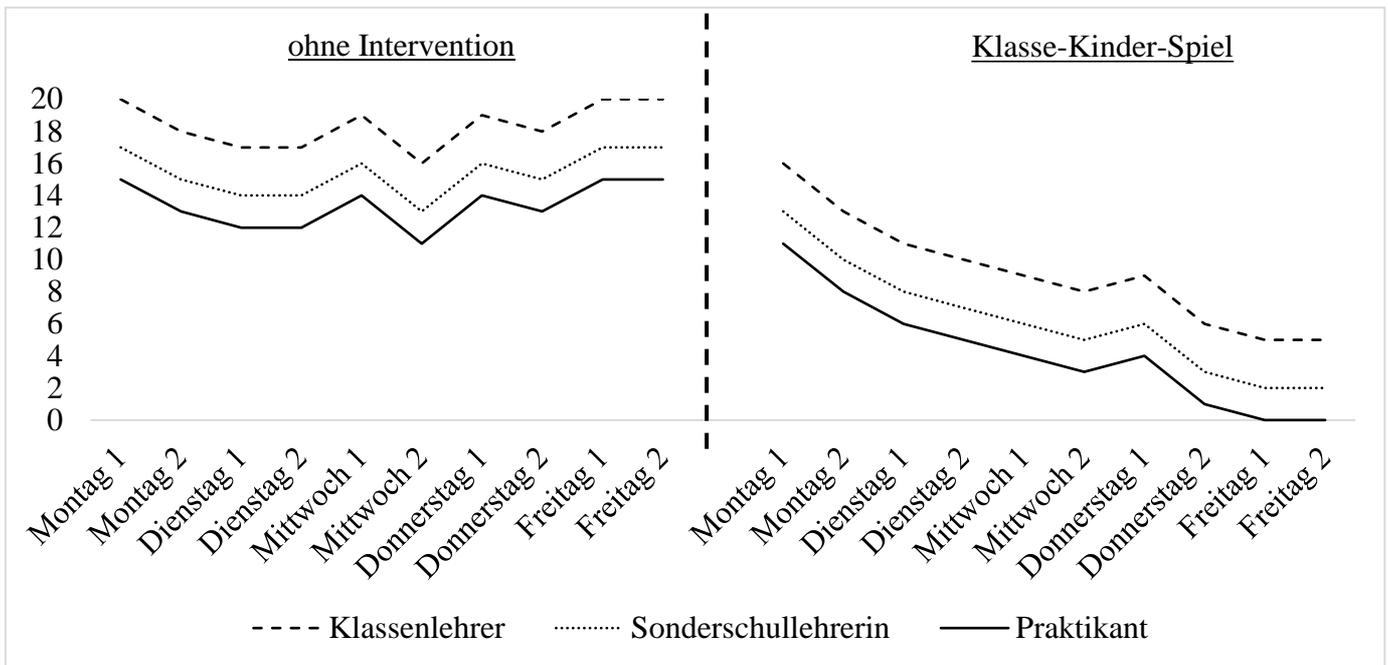


Abbildung 3. Ausmaß des störenden Verhaltens von Lisa in Gruppenarbeitsphasen

Klassenlehrer zu der Entscheidung führt, die Förderung beizubehalten.

Über die konkrete Umsetzung von DBR zur Beurteilung des Fördererfolgs hinaus macht Abbildung 3 noch einen weiteren interessanten Befund deutlich: alle drei Beurteilenden (Klassenlehrkraft, Sonderschullehrkraft, Praktikant) unterscheiden sich zu 100% in ihren absoluten Werten. In diesem Fall liegt demnach eine Beurteilerübereinstimmung von 0% vor und die Befunde sind numerisch variant (Huber & Rietz, 2015a). Interessanterweise sind die Verhaltenstrends der Beurteilenden jedoch identisch, d. h. die Verhaltensunterschiede von Lisa zwischen zwei Messzeitpunkten werden identisch beurteilt. In diesem Fall liegt eine Beurteilerreliabilität von 100% vor und die Befunde sind strukturell invariant (Huber & Rietz, 2015a). Diese Befunde machen deutlich, dass sowohl die beurteilenden Personen als auch

die Zeitpunkte der Beurteilungen einen maßgeblichen Einfluss auf die Ergebnisse – und damit auch auf die Testgüte – der Verhaltensverlaufsdiagnostik haben.

### Personen- und zeitbezogene Einflüsse auf die Messgenauigkeit von DBR

Die Befundlage zur psychometrischen Qualität von DBR ist zum einen noch recht dünn und zum anderen teilweise widersprüchlich. Dies liegt insbesondere daran, dass a) die Methode erst seit ca. 15 Jahren erforscht wird und b) die Durchführung sehr flexibel ist und damit eine Vielzahl an Bedingungen existiert, unter denen DBR durchgeführt werden kann. Aus letzterem Grund werden in psychometrischen Studien vor allem die spezifischen Messbedingungen in den Blick genommen und es wird überprüft, wie sich die Messgenauigkeit unter unterschiedlichen

Bedingungen (z. B. unterschiedliche Rater, Items, Messzeitpunkte) verhält (Christ et al., 2009). Insbesondere personen- und zeitbezogene Faktoren beeinflussen dabei die Messgenauigkeit von DBR. So zeigte sich in einer Studie, dass Lehrkräfte – verglichen mit geschulten externen Beobachter/innen – das Verhalten der gleichen SuS sehr unterschiedlich einschätzen (Chafouleas et al., 2010a). Ähnliche Befunde wurden in weiteren Feldstudien erzielt (Casale, Volpe, Briesch, Hennemann, & Grosche, under review; Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007).

Mit Blick auf die zeitbezogenen Faktoren, wie z. B. die Schultage, konnten ebenfalls moderate Einflüsse auf die Messgenauigkeit gezeigt werden (Briesch, Chafouleas, & Riley-Tillman, 2010; Casale et al., under review). Diese Ergebnisse bedeuten, dass die Messergebnisse unterschiedlich ausfallen je nachdem a) wer die Ratings macht und b) wann die Ratings gemacht werden. Für die praktische Umsetzung sind Unterschiede zwischen den absoluten Ratings verschiedener Personen erst einmal zu vernachlässigen, solange die gleichen Trends erkannt werden (Huber & Rietz, 2015a). Der Einfluss der zeitbezogenen Faktoren ist hingegen ein praktisch relevanter, da er die Frage impliziert, wie viele Messzeitpunkte für zuverlässige pädagogisch-praktische Entscheidungen zu Rate gezogen werden sollten.

Um den Einfluss der zeitbezogenen Faktoren zu verringern und eine möglichst zuverlässige Datengrundlage für pädagogisch-praktische

Entscheidungen zu schaffen, ist die Frage nach der Anzahl benötigter Messzeitpunkte (z. B. Schultage) in der Praxis maßgeblich. Dieser Frage wurde in bisherigen Studien häufig nachgegangen. Die Befundlage diesbezüglich ist jedoch unklar: Je nachdem, welches Verhalten beurteilt wird (z. B. störendes Verhalten, lernförderliches Verhalten) und je nachdem, wer das Verhalten beurteilt (z. B. Psycholog/innen, Lehrkräfte) erwiesen sich zwischen 4 und 20 Schultage als notwendig für zuverlässige Messungen (Briesch et al., 2010; Casale et al., under review; Chafouleas et al., 2010b). Dies bedeutet, dass abhängig von der Beurteilungsdimension sowie der Expertise der Rater/innen innerhalb von einer bis vier Schulwochen eine zuverlässige Datengrundlage für pädagogisch-praktische Entscheidungen erzielt werden kann.

## Zusammenfassendes Fazit

Die Verhaltensverlaufsdiagnostik eignet sich zur Überprüfung des Fördererfolgs einzelner SuS. Sie leistet damit einen wesentlichen Beitrag zur Umsetzung einer evidenzbasierten Praxis. Das Direct Behavior Rating stellt eine Hybridform aus Verhaltensbeurteilungen mit Ratingskalen sowie systematisch-direkten Verhaltensbeobachtungen dar und hat sich für die Zwecke der Verhaltensverlaufsdiagnostik als geeignet erwiesen. Allerdings können die beurteilenden Personen sowie die Tage, an denen beurteilt wird, einen Einfluss auf die Messungen haben. Dies sollte bei der praktischen Umsetzung unbedingt berücksichtigt werden.

Wenngleich es bereits einige Studien zur Testgüte von DBR gibt, ist die Überprüfung dieser Methode noch nicht ausgereift. Zukünftige Aufgaben sind weitere Studien zur Überprüfung der Fragen „Wie gut können DBR Verhaltensänderungen abbilden?“ (Veränderungssensitivität)

oder „Wie gut können DBR ein bestimmtes Verhalten [generell] abbilden?“ (Kriteriumsvalidität) sowie Untersuchungen zu spezifischen Einflussgrößen (z. B. Rater/innen-training, Expertise der Lehrkraft).

## Literaturverzeichnis

- 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*, 66(3), 194–203.
- 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(3), 408–421.
- Casale, G., Hennemann, T., & Grosche, M. (2015). Zum Beitrag der Verlaufsdiagnostik für eine evidenzbasierte sonderpädagogische Praxis am Beispiel des Förderschwerpunkts der emotionalen und sozialen Entwicklung. *Zeitschrift für Heilpädagogik*, 66(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(1), 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(3), 258–268.
- Casale, G., Volpe, R. J., Briesch, A. M., Hennemann, T., & Grosche, M. (under review). Dependability of Direct Behavior Rating Single- and Multi-Item Scales. *Assessment for Effective Intervention*.
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010a). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, 48(3), 219–246.
- Chafouleas, S. M., Briesch, A. M., Riley-Tillman, T. C., Christ, T. J., Black, A. C., & Kilgus, S. P. (2010b). An investigation of the generalizability and dependability of Direct Behavior Rating Single Item Scales (DBR-SIS) to measure academic engagement and disruptive behavior of middle school students. *Journal of School Psychology*, 48(3), 219–246.
- Chafouleas, S. M., Christ, T. J., Riley-Tillman, T. C., Briesch, A. M., & Chanese, J. A. M. (2007). Generalizability and Dependability of Direct Behavior Ratings to Assess Social Behavior of Preschoolers. *School Psychology Review*, 36(1), 63–79.

- 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(4), 201–213.
- Costello, E. J., Copeland, W., & Angold, A. (2011). Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults?: Trends in psychopathology across the adolescent years. *Journal of Child Psychology and Psychiatry*, 52(10), 1015–1025.
- Diehl, K. (2012). IEL-1: Inventar zur Erfassung der Lesekompetenz im 1. Schuljahr – ein curriculumbasiertes Verfahren zur Abbildung des Lernfortschritts. Göttingen: Hogrefe.
- Forness, S. R., Kim, J., & Walker, H. M. (2012). Prevalence of Students with EBD: Impact on General Education. *Beyond Behavior*, 21(2), 3–10.
- Grosche, M. (2014). Fördermaßnahmen im Prozess überprüfen. In T. Bohl, A. Feindt, B. Lütje-Klose, M. Trautmann, & B. Wischer (Hrsg.), *Fördern* (S. 113–115). Velber: Friedrich.
- Haller, A.-C., Klasen, F., Petermann, F., Barkmann, C., Otto, C., Schlack, R., & Ravens-Sieberer, U. (2016). Langzeitfolgen externalisierender Verhaltensauffälligkeiten: Ergebnisse der BELLA-Kohortenstudie. *Kindheit und Entwicklung*, 25(1), 31–40.
- Hillenbrand, C. (2015). Evidenzbasierung sonderpädagogischer Praxis: Widerspruch oder Gelingensbedingung? *Zeitschrift Für Heilpädagogik*, 66(7), 312–324.
- Hillenbrand, C., & Pütz, K. (2008). *KlasseKinderSpiel: Spielerisch Verhaltensregeln lernen*. Hamburg: edition Körber-Stiftung.
- Hölling, H., Schlack, R., Petermann, F., Ravens-Sieberer, U., Mauz, E., & Group, K. S. (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(7), 807–819.
- Huber, C. (2009). Gemeinsam einsam? Empirische Befunde und praxisrelevante Ableitungen zur sozialen Integration von Schülern mit Sonderpädagogischem Förderbedarf im Gemeinsamen Unterricht. *Zeitschrift für Heilpädagogik*, 60(7), 242–248.
- Huber, C., & Grosche, M. (2012). Das response-to-intervention-Modell als Grundlage für einen inklusiven Paradigmenwechsel in der Sonderpädagogik. *Zeitschrift für Heilpädagogik*, 63(8), 312–322.
- Huber, C., & Rietz, C. (2015a). Behavior Assessment Using Direct Behavior Rating (DBR) - A Study on the Criterion Validity of DBR Single-Item-Scales. *Insights on Learning Disabilities*, 12(1), 73–90.
- Huber, C., & Rietz, C. (2015b). Direct Behavior Rating (DBR) als Methode zur Verhaltensverlaufdiagnostik in der Schule: Ein systematisches Review von Methodenstudien. *Empirische Sonderpädagogik*, 7(2), 75–98.
- Krull, J., Wilbert, J., & Hennemann, T. (2014). The Social and Emotional Situation of First Graders with Classroom Behavior Problems and Classroom Learning Difficulties in Inclusive Classes. *Learning Disabilities - A Contemporary Journal*, 12(2), 169–190.
- Makarova, E., Herzog, W., & Schönbächler, M.-T. (2014). Wahrnehmung und Interpretation von Unterrichtsstörungen aus Schülerperspektive sowie aus Sicht der Lehrpersonen. *Psychologie in Erziehung Und Unterricht*, 61(2), 127.

- Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically Derived Subtypes of Child Academic and Behavior Problems: Co-Occurrence and Distal Outcomes. *Journal of Abnormal Child Psychology*, 36(5), 759–770.
- Strathmann, A. & Klauer, K. J. (2012). LVD-M 2-4: Lernverlaufsdiagnostik – Mathematik für zweite bis vierte Klassen. Göttingen: Hogrefe.
- Walter, J. (2014). Lernfortschrittsdiagnostik Lesen (LDL) und Verlaufsdiagnostik sinnerfassenden Lesens (VSL): Zwei Verfahren als Instrumente einer formativ orientierten Lesediagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 165-202). Göttingen: Hogrefe.
- Wilbert, J. (2014). Instrumente zur Lernverlaufsmessung: Gütekriterien und Auswertungsherausforderungen. In M. Hasselhorn, W. Schneider, & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 281–308). Göttingen: Hogrefe.