

Screening von Expansivem Problemverhalten bei Kindern im Kindergartenalter für Eltern und Erzieherinnen (PEP-Screen)

Julia Plück, Christopher Hautmann, Gabriele Brix, Inez Freund-Braier, Kurt Hahlweg und Manfred Döpfner

Zusammenfassung. Indizierte Präventionsmaßnahmen erfordern Screening-Instrumente, die Frühformen, hier expansiven Problemverhaltens (Unaufmerksamkeit, Überaktivität, Impulsivität, Aggressivität), mit möglichst geringem Aufwand erfassen. Zu diesem Zweck wurden 13 Items des *Elternfragebogens über das Verhalten von Kindern und Jugendlichen* (CBCL/4-18) auf der Basis verschiedener Kriterien (Trennschärfindeizes, Faktorenladungen, Itemschwierigkeit, klinische Relevanz im Expertenurteil) für einen Kurzfragebogen ausgewählt, der im Schwerpunkt expansive Verhaltensauffälligkeiten bei Kindergartenkindern im Alter von drei bis sechs Jahren auf der Basis einer Beurteiler übergreifenden Skala aus Eltern- und Erzieherurteil (EL/ER) erfassen sollte. Die Ergebnisse zur Skala *Expansives Verhalten* mittels explorativer Faktorenanalysen an einer kleinen Ausgangsstichprobe ($N_{ER} = 597/N_{EL} = 424$) konnten mit einem konfirmatorischen Vorgehen anhand einer umfangreicheren Replikationsstichprobe ($N_{ER} = 2248/N_{EL} = 1702$) bestätigt und an einer weiteren Stichprobe ($N_{ER} = 599/N_{EL} = 521$) aus einem anderen Erhebungsraum überprüft werden. In der anhand des Prozentrangs 85 der beurteilerübergreifenden Skala (7 Items) indizierten Gruppe konnte bei 20 % der Kinder in der nachfolgenden intensiveren Untersuchung eine Diagnose nach DSM-IV aus dem entsprechenden Bereich vergeben werden.

Schlüsselwörter: indizierte Prävention, expansives Problemverhalten, Screening, Fragebogen

Screening of externalising behavior in preschool children for parents and educators

Abstract. Appropriate indication for prevention needs a screening for early signs of – in this case – externalising behavior (inattentiveness, hyperactivity, impulsivity, and aggression). For this reason, 13 items of the *Child Behavior Checklist (CBCL/4-18)* were selected by empirical (item total correlation in a representative sample, factor analyses) and clinical (expert rating) decision, and included in a short questionnaire for parents and kindergarten teachers (P/T) of children aged 3 to 6 years, focussing on the cross-informant view of externalising behavior. Findings of exploratory factor analysis based on a small sample ($N_T = 597/N_P = 424$) generated in the city of Cologne could be replicated by confirmatory analyses including a more extensive sample from Cologne ($N_T = 2248/N_P = 1702$) and another sample from the city of Brunswick ($N_T = 599/N_P = 521$) and confirmed a 7-item-scale of *externalising behavior*. 20 % of the children within the sample indicated by the 85th percentile of the cross-informant externalising scale could be classified as showing disruptive behavior following DSM-IV criteria.

Key words: indicated prevention, externalising behavior, screening, short questionnaire

Bei der Erfassung expansiven Verhaltens als Sammelbegriff für Unaufmerksamkeit, Überaktivität, Impulsivität und Aggressivität gerade im frühen Kindesalter reicht dieses in einem dimensionalen Verständnis von angemessenem Verhalten bis hin zur schweren Störung im Sinne einer Diagnose (z. B. Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS), Störungen des Sozialverhaltens). Die Prä-

valenzraten entsprechender Diagnosen reichen in Feldstichproben bedingt z. B. durch unterschiedliches Alter und diagnostische Methoden und Kriterien von ca. 5 % bis 20 % (Ihle et al., 2004; Miller et al., 2002). Externalisierende Probleme werden allgemein als beste Prädiktoren für zukünftiges oppositionelles Verhalten und dissoziale Störungen betrachtet und in unbehandelten Populationen entwickeln etwa 50 % der Kinder mit deutlichen externalisierenden Verhaltensproblemen dauerhafte oppositionelle Verhaltensprobleme (Bennett et al., 1999).

Diese Studie wurde von der Deutschen Forschungsgemeinschaft finanziell gefördert (DFG, Projekt DO 620/ 2-1; Wirksamkeit indizierter Prävention bei external auffälligen Vorschulkindern) und fand in Kooperation mit der Universität Braunschweig statt (Projektleiter Kurt Hahlweg, DFG, Projekt HA 1400/14/1-4).

Auf diesen Grundlagen zieht die Forderung nach präventiven Maßnahmen hinsichtlich expansiven Problem-

verhaltens mit Blick auf eine möglichst zielgenaue Bündelung der Ressourcen die Idee der Konzentration auf solche Kinder nach sich, die frühe Zeichen oder Symptome einer Störung aufweisen, ohne (zwingend) alle Kriterien einer psychischen Störung zu erfüllen (indizierte Prävention). Dies wirft die Frage auf, wie entsprechende Frühformen schon im Kindergartenalter mit möglichst geringem Aufwand erfasst werden können.

Am häufigsten werden Breitbandverfahren wie der Strengths and Difficulties Questionnaire (SDQ) von Goodman (1997) oder die diversen Formen der Child Behavior Checklist (CBCL) von Achenbach (z.B. Achenbach & Rescorla, 2000) zum „Screening“ als erstem Schritt in der klinischen Diagnostik bei Inanspruchnahmestichproben eingesetzt und in diesem Rahmen auf ihre differenzielle Kraft untersucht (Mullick & Goodman, 2001; Rescorla, 2005). Verschiedene Studien mit der CBCL/4-18 unterstreichen die prädiktive Eignung der Skalen *Aufmerksamkeitsprobleme* bzw. *Aggressives Verhalten* für die Diagnose einer Aufmerksamkeitsdefizits-/Hyperaktivitätsstörung (ADHS) bzw. einer Störung des Sozialverhaltens (Störung mit oppositionellem Trotzverhalten nach DSM-IV) (z.B. Hudziak et al., 2004; Lampert et al., 2004; Kasius et al., 1997). Da in klinischen Stichproben jedoch von deutlich höheren Prävalenzraten als im Feld auszugehen ist, können ihre Ergebnisse nur als Hinweis darauf verstanden werden, dass diese Instrumente eine geeignete Ausgangsbasis auch für ein Screening in Feldstichproben darstellen. An einer Stichprobe von Grundschulkindern konnten Van Lier und Mitarbeiter (2003) unter Einsatz der CBCL feststellen, dass für ein möglichst genaues Screening hinsichtlich späterer Auffälligkeiten in den Bereichen *ADHS*, *ADHS mit oppositionellem Trotzverhalten* und *Störungen des Sozialverhaltens* Symptome oppositionellen, aufmerksamkeitsgestörten Verhaltens und gestörten Sozialverhaltens einbezogen werden mussten. Entsprechende Untersuchungen unter Einsatz des SDQ (z.B. Goodman et al., 2003; Maniadaki et al., 2005) belegen die Eignung auch dieses Instruments für entsprechende Screening-Zwecke an Feldstichproben.

Allgemeine Kriterien für die Bewertung von Screening-Prozeduren formulierten bereits Wilson und Jungner (1968) für die WHO. Diese reichen vom Vorliegen von Erkenntnissen zum betreffenden Krankheitsbild (Bedeutbarkeit, Erkennbarkeit von Frühstadien, Verständnis des natürlichen Krankheitsverlaufs) über das entsprechende Untersuchungsverfahren (Eignung, Akzeptanz in der Bevölkerung, Möglichkeit kontinuierliche Untersuchungen) und Behandlungsmöglichkeiten (akzeptierte Behandlungsprogramme, Verfügbarkeit von Diagnose- und Behandlungsangeboten, Konsens über Kriterien für die Störungsdefinition) bis hin zu Kosten-Nutzen-Abwägungen. Im Rahmen der Früherkennung gegenüber dem Einsatz in der Versorgung gewinnt die positive prädiktive Kraft (PPK) eines Fragebogens, also die Wahrscheinlichkeit, dass ein als „auffällig“ klassifizierter Fall bei genauerer Betrachtung auch tatsächlich eine Diagnose erhält nach Muir Gray (1994) größere Bedeutung gegenüber anderen Bewertungskriterien (z.B. Sensitivität, Spezifität).

Außerdem hängen diese neben dem Screening-Instrument selbst von der Stichprobe (z.B. unterschiedliche Prävalenzraten in Klinik- und in Feldstichproben), dem Alter, dem Abstand zwischen Screening und Kriterium, und eben diesem Kriterium ab und sind daher über verschiedene Studien hinweg schlecht vergleichbar.

Zur Indizierung werden in den meisten Studien einfache Klassifikationsstrategien verwendet, z.B. eine Aufteilung der Population anhand eines mehr oder weniger frei gewählten Cut-Offs hinsichtlich eines einzelnen Kriteriums, meist externalisierender Verhaltenssymptome. Bennett und Mitarbeiter (1998, 1999) untersuchten die Eignung verschiedener Vorgehensweisen zur Ermittlung von frühen Risikofaktoren für oppositionelle Probleme und als Indikatoren für gezielte Präventionsmaßnahmen. Für externalisierende Verhaltensprobleme stellten sie übergreifend eine Sensitivität der verschiedenen Strategien von unter 50 % fest. Sie räumen allerdings ein, dass selbst wenn man nur externalisierende Symptome zur Indizierung nutzt, die Rate der Versorgung von 20 % (klinische Inanspruchnahme) auf über 40 % gesteigert würde. Im Vergleich zu universeller Prävention sei indizierte Prävention daher selbst bei schwächerer Klassifikationsgüte des Screening-Verfahrens immer noch ökonomischer, die ja bei einer Prävalenzrate von 15 % immerhin 85 % „umsonst“ behandeln würde.

In der Studie der Conduct Problem Prevention Research Group (Lochmann & CPPRG, 1995; Hill et al., 2004) an Schulkindern wurde zur Indizierung eine „multiple-gating“-Strategie verwendet. Die Rekrutierung für die Evaluationsstudie zum FAST-TRACK-Programm erfolgte in belasteten Umgebungen (selektiv). Hier wurden anhand des Erzieherurteils mit einer Itemauswahl aus einem etablierten Fragebogen (TOCA-R) (Werthamer-Larsson et al., 1991) die Auffälligsten ausgewählt, deren Verhalten daraufhin von ihren Eltern anhand eines eigens für die Studie konzipierten Instruments aus Items der CBCL und der Revised Problem Behavior Checklist (Lochman et al., 1995) beurteilt wurde. Als Kriterium zur Überprüfung der Klassifikationsgüte dienten die übergeordneten Skalen der CBCL. In der Kontrollgruppe konnte eine PPK von 86 % beim 3-Jahres-Follow-Up erreicht werden. Es zeigten sich keine Vorteile gegenüber dem früher erprobten „1-Gate“-Modell (Reid, 1993). Da Kinder mit situationsübergreifenden Problemen ein größeres Risiko für spätere Probleme tragen (Loeber, 1990) sollten durch den Einbezug verschiedener Informationsquellen unterschiedliche „settings“ repräsentiert sein. So trüge man auch den erfahrungsgemäß insbesondere im Vorschulalter geringen Korrelationen Rechnung (Verhulst & Akkerhuis, 1989; Cai et al., 2004; Rescorla, 2005), die sowohl durch tatsächliche Abweichungen, als auch durch Wahrnehmungsverzerrungen sowie durch Situationsspezifität des kindlichen Verhaltens bedingt sein könnten (Döpfner & Lehmkuhl, 1997). Lochman und Mitarbeiter (1995) verweisen darauf, dass im Idealfall die Quellen für das Screening-Kriterium auch den Ansatzpunkten für die nachgeschaltete Interventionsmaßnahme entsprechen sollten.

Die vorliegende Studie dient der Entwicklung eines Screening-Fragebogens für expansives Problemverhalten (*PEP-Screen*) bei Kindergartenkindern unter Einbezug der Eltern- und der Erzieherperspektive. Sie ist Teil der umfangreicheren DFG-Studie (Projektleiter: Manfred Döpfner, DO 620/ 2-1: Wirksamkeit indizierter Prävention bei external auffälligen Vorschulkindern) zur Evaluation des Präventionsprogramms für Expansives Problemverhalten (PEP) (Plück et al., 2006) in der das Instrument zur Identifikation der Zielpopulation eingesetzt werden sollte. Erste Ergebnisse der Evaluation der indizierten Präventionsmaßnahme finden sich bei Hanisch et al. (2006).

Methoden

Stichproben

Für die Überprüfung des *PEP-Screen* in der vorliegenden Studie wurden drei Stichproben genutzt, die im Rahmen der oben genannten DFG-Studie in Köln und Braunschweig erhoben wurden. Dabei kann bei der angegebenen Anzahl von Erzieherfragebögen von einer Vollerhebung der Zielstichprobe 3- bis 6-jähriger Kinder in den jeweiligen Kindergärten ausgegangen werden. In der zentralen Tendenz zeigten sich keine systematischen Unterschiede im Erzieherurteil zwischen Kindern mit und ohne Vorliegen eines Elternfragebogens hinsichtlich der Zieldimension *Expansives Verhalten* und der *Gesamtskala*. Aus zusätzlichen jedoch nur lückenhaft vorliegenden Angaben der Erzieherinnen ließ sich auf Sprachprobleme der Eltern als häufigsten Grund für eine Nicht-Teilnahme an der Befragung schließen. Im Rahmen der Evaluationsstudie eines in deutscher Sprache konzipierten Präventionsprogramms war dies allerdings zu erwarten und wirft für die Verbreitung in speziellen Risikogruppen sicher weitere Fragen auf. Im Einzelnen wurden folgende Stichproben untersucht:

1. Analysestichprobe (Köln I):

In der ersten Phase der ersten Erhebungswelle im Jahr 2001 wurden in Köln Erzieherurteile über $N = 597$ und Elternurteile über $N = 424$ Kinder aus 17 von 20 angesprochenen städtischen Tageseinrichtungen für Kinder (Juli bis August 2001) erreicht: 52,4% Jungen, Altermittelwert 4,2 Jahre ($SD = 0.76$).

2. Replikationsstichprobe A – Köln II:

Die Erhebungen von Welle1 wurden bis Oktober 2001 um Neuzugänge in den bereits rekrutierten Einrichtungen ergänzt und im Jahr 2002 wurde die umfangreichere Welle2 in weiteren 48 Einrichtungen (Juni bis September 2002) erhoben. Am Ende der Gesamterhebungen lagen $N = 2845$ Erzieherurteile und $N = 2133$ Elternurteile über Kölner Kinder aus insgesamt 62 von 68 angesprochenen städtischen Tageseinrichtungen vor. Abzüglich der Stichprobe Köln I stand also eine Replikationsstichprobe von $N = 2248$ Erzieher- und $N = 1702$ Elternurteilen (75,1%) zur Verfügung. Diese Stichprobe wies hinsichtlich der Mehrheit der erfassten soziodemographischen Variablen keine signifi-

kanten Abweichungen von Köln I auf, war jedoch durch eine geringere Schulbildung der Mutter (Rangtest: $p \leq .05$) und einen höheren Grad an sozialer Belastung im Wohnumfeld (Rangtest: $p \leq .001$) gekennzeichnet. Darüber hinaus wiesen die Kinder ein signifikant geringeres Alter auf ($AM (SD) = 4.04 (.88)$ Jahre; t -Test: $p \leq .01$). Diese Unterschiede sind plausibel, da sich die ergänzend erfassten jüngeren Kinder, die nach den Sommerferien neu in die Einrichtungen kamen, aus beiden Jahren in der Stichprobe Köln II befinden. Zur Nutzung einer möglichst breiten Datengrundlage scheint eine Anpassung der Replikationsstichprobe A hinsichtlich der festgestellten Unterschiede jedoch nicht sinnvoll. Die beiden Stichproben aus Köln bilden zusammen die Stichprobe *Köln – Gesamt*.

3. Replikationsstichprobe B:

In Braunschweig wurden nach den Sommerferien des Jahres 2002 in insgesamt 10 von 11 angesprochenen städtischen Tageseinrichtungen für Kinder Erzieherurteile über $N = 599$ Kinder eingeholt. Der Anteil der Eltern, die sich ebenfalls an der Studie beteiligten, lag in Braunschweig mit $N = 521$ (87,1%) signifikant höher als in der Replikationsstichprobe A (Köln II) (χ^2 -Test: $p \leq .001$) und Köln – Gesamt (75,1%; χ^2 -Test: $p \leq .001$), was insbesondere auf einen geringeren Ausländeranteil in Braunschweig zurückzuführen sein dürfte. Die Geschlechtsverteilung sah mit 49,8% Jungen ähnlich ausgefallen aus wie in Köln, jedoch waren die Kinder signifikant älter ($AM (SD) = 4.65 (1.41)$ Jahre; t -Test: $p \leq .001$). Insgesamt zeigte sich ein Trend zu höheren Werten auf den Variablen zur elterlichen Schul- und Berufsausbildung im Vergleich zu Köln, der außer für die Schulbildung der Mutter, sowohl für die des Vaters (Rangtest: $p \leq .05$) und für die berufliche Bildung von Mutter und Vater (Rangtests: $p \leq .001$) signifikant ausfiel. Diese Unterschiede geben Hinweise auf eine voneinander abweichende Sozialstruktur in beiden Städten, können aber zumindest zusätzlich auch durch Unterschiede in der Stichprobenziehung z. B. hinsichtlich Auswahl der Einrichtungen und des Erhebungszeitpunkts bedingt sein. Sollte die faktorielle Struktur des Fragebogens sich trotz dieser Unterschiede jedoch bestätigen lassen, wäre dies – wie auch schon innerhalb der Kölner Stichprobe – als Zeichen der Robustheit des Fragebogens zu werten, so dass keine Maßnahmen zur Anpassung der Stichproben unternommen wurden.

Darüber hinaus lagen differenzierte diagnostische Informationen für die Kinder der indizierten Teilstichprobe aus Köln vor, deren Eltern mit einer Kontaktaufnahme einverstanden und zur Teilnahme am „Prätest“ (umfangreiche Fragebogenerhebungen bei Eltern und Erzieherin zzgl. Hausbesuch mit umfangreichen Untersuchungen) im Rahmen der o. g. Evaluationsstudie etwa zwei bis drei Monate nach dem Screening bereit waren ($n = 155$).

Instrumente

Die Zusammenstellung einer überschaubaren Anzahl von Items für den PEP-Screen für Erzieherinnen (ER) und El-

tern (EL) zur Erfassung der Zielgruppe für das indizierte *Präventionsprogramm für Expansives Problemverhalten* (PEP) in der betreffenden Altersgruppe (3 bis 6 Jahre) sollte expansive sowie einige komorbide Auffälligkeiten berücksichtigen. Zur Kompensation der Sicht von Eltern und Erzieherin wurde als Indizierungskriterium für eine Gesamtskala *Expansives Verhalten* (Aufsummierung der entsprechenden Items aus beiden Perspektiven) entschieden, so dass sie mit gleichem Gewicht eingehen. Dem präventiven Ansatz folgend sollten auch solche Kinder indiziert werden können, bei denen die Schwierigkeiten noch nicht in höchstem Maße übergreifend ausgeprägt waren. Dies wäre mit einer „Schnittmenge“ aus den pro Perspektive unter Anwendung z. B. des gleichen Prozentrangs als Cut-Off übereinstimmend als „auffällig“ Klassifizierten nicht gegeben gewesen. Die beiden in der Forschung am weitesten verbreiteten Instrumente für eine breitbandige Verhaltensdiagnostik schienen für ein gezieltes Screening expansiven Verhaltens entweder zu umfangreich (Fragebogen für Eltern von Klein- und Vorschulkindern/CBCL1 ½–5 bzw. Fragebogen für Erzieherinnen von Klein- und Vorschulkindern/C-TRF; Arbeitsgruppe Deutsche Child Behavior Checklist, 2000a, 2000b) oder bildeten nach klinischer Abwägung keine ausreichende Vielfalt ab (Strengths and Difficulties Questionnaire/SDQ; Goodman, 1997). Für eine geeignete Itemauswahl nach empirischen Kriterien konnte auf den Itempool der CBCL/4–18 zurückgegriffen werden, für den die Teilstichprobe der Vorschulkinder ($N = 455$) der deutschen bundesweit repräsentativen Normstichprobe (Döpfner et al., 1998) zur Verfügung stand. Wie schon in der Studie der CPPRG (Lochman & CPPRG, 1995; Hill et al., 2004) wurde eine Auswahl – hier von 13 Items – getroffen, die in der genannten Stichprobe auf den jeweiligen Skalen vergleichsweise hohe Trennschärfen aufwiesen und im Konsens von drei klinisch-psychologischen Fachleuten als bedeutsam bewertet wurden. Die Skalierung der Items wurde analog CBCL beibehalten und reichte von „0 = nicht“ über „1 = etwas/manchmal“ bis „2 = genau/häufig zutreffend“. Für den Kernbereich expansiven Verhaltens wurden aus den CBCL-Skalen *Aufmerksamkeitsprobleme*, *Dissoziales Verhalten* und *Aggressives Verhalten* insgesamt 8 Items ausgewählt (3. Streitet oder widerspricht viel, 8. Kann sich nicht konzentrieren, kann nicht lange aufpassen, 10. Kann nicht stillsitzen, ist unruhig oder überaktiv, 19. Verlangt viel Beachtung, 21. Macht Sachen kaputt, die anderen gehören, 41. Ist impulsiv oder handelt ohne zu überlegen, 57. Greift andere körperlich an, 95. Hat Wutausbrüche oder hitziges Temperament). Die interne Konsistenz der so gebildeten Skala *Expansives Verhalten* fiel in dieser Stichprobe mit $\alpha = .77$ zufriedenstellend aus und es konnten Item-Trennschärfen von $r_{it} = .32$ (Item 21) bis $r_{it} = .59$ (Item 3) ermittelt werden. Die Auswahl der Items zum expansiven Verhalten schien nach diesen Ergebnissen, die jedoch nur auf Informationen aus dem Elternurteil basieren, vorläufig gerechtfertigt.

Darüber hinaus wurden zur Erfassung komorbider internalisierender Auffälligkeiten aus den CBCL-Skalen *Sozialer Rückzug*, *Ängstlich/depressiv* und *Soziale Probleme* weitere 4 Items sowie ein Item aus *Körperliche Be-*

schwerden ausgewählt (11. Klammert sich an Erwachsene oder ist zu abhängig, 38. Wird viel gehänselt, 50. Ist zu furchtsam oder ängstlich, 103. Ist unglücklich, traurig oder niedergeschlagen, 56a. Sagt häufig, dass ihm etwas wehtut (ohne bekannte körperliche Ursache)). In der gemeinsamen Analyse wurden für diese Items Trennschärfen von $r_{it} = .26$ (Item 103) bis $r_{it} = .36$ (Item 11) ermittelt. Für das Item 56a (Schmerzen) konnte lediglich eine Trennschärfe von $r_{it} = .03$ festgestellt werden. Hier entschied die klinische Erwägung (s. o.) über den Einschluss. Die interne Konsistenz der so gebildeten Skala *Internalisierendes Verhalten* lag mit $\alpha = .45$ erwartungsgemäß (auch unter Verzicht von Item 56a, $\alpha = .49$) nicht im zufriedenstellenden Bereich. Aber auch Analysen mit einer veränderten Itemauswahl ließen keine konsistente eigenständige Skala mit relativ wenigen internalisierenden Items erwarten. Die zufriedenstellende interne Konsistenz der *Gesamtskala* ($\alpha = .80$) aus allen 13 Items verhielt zumindest orientierende Informationen über die psychische Verfassung der Kinder unter Einbezug emotionaler Probleme. Symptome aus diesem Bereich treten außerdem nicht selten auch komorbid zu expansiven Problemen auf und sollten gerade im präventiven Bereich mit berücksichtigt werden. Darüber hinaus sprach gegen eine Reduktion auf die Erfassung rein expansiven Verhaltens auch die Befürchtung einer dann geringeren Bereitschaft bei den zu befragenden Eltern und Erzieherinnen, sich an diesem ersten Schritt der Studie zu beteiligen. Schon bei der Vorstellung der Studie in den Kindergärten und der Werbung um Teilnahme wurde dann sowohl von Erzieherinnen wie von Elternvertretern dieser Ansatz positiv gewürdigt.

Zur Überprüfung des *PEP-Screen* wurden verschiedene Skalen und Itemgruppen aus weiteren Instrumenten der breitbandigen und der störungsspezifischen Diagnostik von Verhaltensproblemen im Kindesalter herangezogen, die für eine Teilstichprobe in zeitlichem Abstand zum Screening im Rahmen des Prätests der Evaluationsstudie bei der indizierten Gruppe zusätzlich erhoben werden konnten: Aus den Fremdbeurteilungsbögen zur Hyperkinetischen Störung und zur Störung des Sozialverhaltens (FBB-HKS bzw. FBB-SSV) aus dem Diagnostik-System für psychische Störungen im Kindes- und Jugendalter nach ICD10 und DSM-IV (DISYPS-KJ; Döpfner & Lehmkuhl, 2000) wurde die HKS-Gesamt-Skala und die Skala Oppositionelles Verhalten herangezogen. Auf der Basis dieser Fragebögen und der entsprechenden Diagnose-Checklisten aus Sicht der Kollegen, die den Hausbesuch durchführten, wurden Diagnosen nach DSM-IV generiert. Aus dem Fragebogen für Eltern bzw. Erzieherinnen von Klein- und Vorschulkindern (CBCL/1½–5, C-TRF/1½–5) (Arbeitsgruppe Deutsche Child Behavior Checklist, 2000a, 2000b; Achenbach & Rescorla, 2000) wurden die Skalen *Externalisierendes Verhalten* und *Internalisierendes Verhalten* verwendet. Zusätzlich wurde auf der Basis der Items der CBCL bzw. der C-TRF/1½–5 die Beurteiler übergreifende Skala *Expansives Verhalten* zu diesem späteren Untersuchungszeitpunkt „rekonstruiert“. An Stelle der in dieser Vorschulalter-Version der Fragebögen (im Gegensatz zu der Ausgangsversion der CBCL/4–18) fehlenden Items „Streitet oder widerspricht viel“ und „Ist impulsiv

oder handelt ohne zu überlegen“ wurden ersatzhalber die Items 20 „Gehorcht nicht“ und 59 „Wechselt sehr schnell von einer Beschäftigung zur nächsten“ eingesetzt, die den erfassten Symptomen inhaltlich am nächsten kamen. Für die Elternskala, Erzieher skala und die Beurteiler übergreifende Skala wurden zufriedenstellende interne Konsistenzen von $\alpha = .75$ bis $\alpha = .76$ ermittelt.

Analysen

An der Analysestichprobe (Köln I) wurden die faktorielle Struktur des Fragebogens mittels explorativer Faktorenanalysen (EFAs: Hauptkomponentenanalyse mit Varimax-Rotation) jeweils für Eltern und Erzieherinnen ermittelt und die entsprechenden internen Konsistenzen geprüft. Auf dieser Basis wurde die Skalenstruktur der beiden parallelen Instrumente für den weiteren Verlauf der Studie festgelegt. Die Skala *Expansives Verhalten* als zentrales Kriterium für die Indizierung der Kinder für eine Teilnahmeempfehlung am *Präventionsprogramm für Expansives Problemverhalten* (PEP), wurde anhand der Replikationsstichprobe A (umfangreicher, gleiche Stadt) durch konfirmatorische Faktorenanalysen überprüft. Im Anschluss wurden zur Feststellung lokaler Unterschiede und der Stabilität der Skala *Expansives Verhalten* über verschiedene Stichproben hinweg im Sinne einer Kreuzvalidierung vergleichende konfirmatorische Analysen zwischen Replikationsstichprobe A und B durchgeführt (Mehrgruppenanalyse). Die Ermittlung der Trennschärfen der Items und der

internen Konsistenzen der Skalen erfolgte in den beiden Gesamtstichproben aus Köln und Braunschweig. Darüber hinaus wurden innerhalb der erreichbaren indizierten Kinder der Kölner Gesamtstichprobe verschiedene Indikatoren für die Stabilität und die Validität der Erfassung expansiven Verhaltens mittels des *PEP-Screen* ermittelt.

Ergebnisse

Explorative Analysen

Bei der explorativen Analyse der Items in der Stichprobe Köln I konnten im Erzieherurteil mit geringen Abweichungen die angestrebten zwei Faktoren (Internal/External) mit einem Eigenwert größer 1 extrahiert werden. Im Elternurteil ergaben sich mit diesem Vorgehen drei Faktoren, wobei die externalisierenden Items sich auf 2 Faktoren verteilten. Es wurde daher als Hauptanalyse eine 2-faktorielle Lösung in beiden Perspektiven vorgegeben. In Tabelle 1 sind die Ladungen der Items auf den beiden vorgegebenen Faktoren aufgeführt.

Es ergab sich sowohl für die Erzieherinnenperspektive (Varianzaufklärung 52 %) als auch für die Elternperspektive (Varianzaufklärung 38 %) weitgehend die erwartete Zuordnung der Items zu den Faktoren. Item 3 zeigte in beiden Perspektiven Doppelladungen, die – entgegen der Erwartung – auf dem Faktor höher ausfielen, auf dem ansonsten die eher internalisierenden Items luden. Diese Verlagerung

Tabelle 1. Explorative Faktorenanalysen (EFAs)¹ über die Items des PEP-Screening-Fragebogens (PEP-Screen) für Eltern und Erzieherinnen an der Ausgangsstichprobe (Köln I)

| | Erzieherinnen n = 592 | | Eltern n = 430 | |
|--|--------------------------|------|-------------------|------|
| | 1 | 2 | 1 | 2 |
| 1. Streitet oder widerspricht viel | .78 | .09 | .60 | .28 |
| 3. Verlangt viel Beachtung | .35 | .66 | .41 | .48 |
| 5. Kann sich nicht konzentrieren, kann nicht lange aufpassen | .67 | .15 | .54 | .28 |
| 6. Ist unruhig überaktiv | .78 | .05 | .58 | .10 |
| 8. Macht Sachen kaputt, die anderen gehören | .76 | -.09 | .59 | -.04 |
| 10. Ist impulsiv oder handelt ohne zu überlegen | .87 | -.02 | .68 | .01 |
| 12. Greift andere körperlich an | .81 | -.12 | .68 | -.01 |
| 13. Hat Wutausbrüche oder hitziges Temperament | .72 | -.01 | .62 | .09 |
| 2. Wird viel gehänselt | .32 | .35 | .14 | .40 |
| 4. Klammert sich an Erwachsene oder ist zu abhängig | .02 | .76 | .01 | .77 |
| 7. Ist zu furchtsam oder ängstlich | -.23 | .62 | -.09 | .64 |
| 9. Ist unglücklich, traurig oder niedergeschlagen | -.05 | .68 | .09 | .54 |
| 11. Sagt häufig, dass ihm etwas weh tut (ohne körperliche Ursache) | .02 | .53 | .09 | .51 |
| Eigenwert | 4.4 | 2.3 | 2.9 | 2.1 |

Anmerkung: ¹ Hauptkomponentenanalyse mit Varimaxrotation.

war besonders deutlich im Urteil der Erzieherinnen der Fall und ließ eine Informanten übergreifende Zuordnung des Items zu einem Faktor nicht zu. Auch für Item 2 rechtfertigten niedrige Doppelladungen im Erzieherurteil keine eindeutige Faktorzuordnung. Die interne Konsistenz der Skala *Expansives Verhalten* profitierte im Erzieherurteil vom Ausschluss des Items 3 ($r_{it} = .33$; von $\alpha = .87$ auf $\alpha = .89$) während im Elternurteil die interne Konsistenz unbedeutend vermindert wurde (von $\alpha = .76$ auf $\alpha = .75$). Auf Grund der vergleichsweise niedrigen Trennschärfe ($r_{it} = .40$) und des Ziels einer itemgleichen Skalenstruktur für beide Perspektiven schien der Ausschluss des Items aus dieser Skala jedoch auch hier gerechtfertigt. Hinsichtlich Item 2 sprach im Erzieherurteil die Trennschärfe von $r_{it} = .20$ und die Verbesserung der internen Konsistenz von $\alpha = .60$ auf $\alpha = .62$ für einen Ausschluss aus der Skala *Internalisierendes Verhalten*. Da das Item auch in der Elternperspektive eine eher geringe Trennschärfe aufwies ($r_{it} = .21$) und die ohnehin geringe interne Konsistenz der Skala ($\alpha = .54$) nicht weiter reduziert wurde, konnte der Ausschluss auch hier vertreten werden. Allerdings signalisieren diese Werte eine insgesamt nur sehr eingeschränkte Verwendbarkeit der Skala *Internalisierendes Verhalten*. In beiden Beurteilerperspektiven zeigten die beiden fraglichen Items auf der *Gesamtskala* zumindest nicht die jeweils geringsten Trennschärfen (ER: $r_{it3} = .50/r_{it2} = .35$; EL: $r_{it3} = .47/r_{it2} = .23$). Die interne Konsistenz der *Gesamtskala* konnte daher durch den Ausschluss der beiden Items nicht erhöht werden und es wurde entschieden, sie im Fragebogen zu belassen und gemeinsam mit den übrigen Items

jedoch nur bei der Berechnung der *Gesamtskala* zu berücksichtigen.

Konfirmatorische Faktorenanalysen an den Replikationsstichproben A und B

Zunächst sollte eine Prüfung des vorgeschlagenen 2-faktoriellen Modells (unter Abzug der Items 2 und 3) in der Replikationsstichprobe A, die sich in den explorativen Analysen zwar allein als faktoriell instabil, in der Gesamtskala jedoch als vertretbar gezeigt hatten, über den Verbleib der internalisierenden Items im Fragebogen entscheiden. Im Erzieherurteil konnte unter Berücksichtigung der vom Auswertungsprogramm LISREL (Jöreskog & Sörbom, 1996) ausgegebenen Modifikationswerte, die eine Kovarianz der Items 5 und 6 innerhalb des Expansiven Faktors nahe legten mit einem CFI = .97 (Comparative Fit Index; Bentler, 1990) und einem SRMR = .097 (Standardized Root Mean Square Residual) gerade noch vertretbare Werte erreicht werden. Im Elternurteil hingegen fiel der CFI-Wert mit 0.89 grenzwertig aus, während der SRMR mit .076 vertretbar schien.

Die weitere Überprüfung des *PEP-Screen* konzentrierte sich jedoch auf die Skala *Expansives Verhalten*, da diese als Indizierungskriterium dienen sollte.

Zunächst wurden in Replikationsstichprobe A für beide Perspektiven (Eltern = EL, Erzieherinnen = ER) einfaktorielle Modelle aufgestellt, bei denen alle Item auf einer la-

Tabelle 2. Faktorladungen¹ der Items der Skala *Expansives Verhalten* in konfirmatorischen Faktorenanalysen² des *PEP-Screen* für Eltern und Erzieherinnen in den Replikationsstichproben A (Köln) und B (Braunschweig; in Klammern); Kennwerte der Modellgüte

| | Erzieherinnen <i>N</i> = 2231 (<i>N</i> = 592) | Eltern <i>N</i> = 1657 (<i>N</i> = 518) |
|--|---|--|
| 1. Streitet oder widerspricht viel | .84 (.84) | .69 (.66) |
| 5. Kann sich nicht konzentrieren, kann nicht lange aufpassen | .70 (.63) | .50 (.55) |
| 6. Ist unruhig überaktiv | .82 (.76) | .55 (.61) |
| 8. Macht Sachen kaputt, die anderen gehören | .92 (.90) | .65 (.74) |
| 10. Ist impulsiv oder handelt ohne zu überlegen | .93 (.94) | .76 (.81) |
| 12. Greift andere körperlich an | .91 (.91) | .69 (.81) |
| 13. Hat Wutausbrüche oder hitziges Temperament | .87 (.87) | .72 (.69) |
| CFI ³ | .99 | .95 |
| χ^2 ; <i>p</i> | 124.02; <i>p</i> ≤ .000 (48.48; <i>p</i> ≤ .000) | 90.44; <i>p</i> ≤ .000 (29.60; <i>p</i> ≤ .000) |
| Modell1: keine Faktorladung invariant (χ^2 ; <i>df</i>) | 172.50; 26 | 120.04; 26 |
| Modell2: alle Faktorladungen invariant (χ^2 ; <i>df</i>) | 176.59; 33 | 132.68; 33 |
| Modellvergleich ⁴ ($\Delta\chi^2$; Δdf ; Signifikanz) | 4.09; 7; n.s. | 12.65; 7; n.s. |

Anmerkungen: ¹Weighted Least Squares, Einfaktorielle Lösung, korrelierende Messfehler von Item 5 und Item 6. ²Einfaktorielle Lösung, korrelierende Messfehler von Item 5 und Item 6, alle Faktorladungen invariant. ³Comparative Fit Index : Index der Modellgüte. ⁴ χ^2 -Test über die Differenz (Δ) der χ^2 -Werte der einzelnen Modelle unter Berücksichtigung der Differenz der Freiheitsgrade.

tenten Variable „Expansives Verhalten“ luden. Unter Berücksichtigung der auch in diesem Modell nahe gelegten Modifikation wegen Kovarianz der Items 5 und 6 ergaben sich als Marker für die Modellgüte $CFI_{ER} = .99$ und $CFI_{EL} = .95$ bzw. $SRMR_{ER} = .068$ und $SRMR_{EL} = .061$. Der χ^2 -Anpassungstest fiel jedoch sowohl für Erzieherinnen ($\chi^2_{ER} (df = 13, N = 2231) = 124.02, p \leq .000$) als auch für Eltern ($\chi^2_{EL} (df = 13, N = 1657) = 90.44; p \leq .000$) an dieser Stichprobe signifikant aus. Demnach wäre beim χ^2 -Anpassungstest die Modellanpassung zwar in Frage gestellt gewesen, wegen der hohen Power bei großen Stichproben besteht beim χ^2 -Test allerdings die Gefahr, ein an sich passendes Modell zu verwerfen (Hu & Bentler, 1995). Auf Grund der guten CFI- und SRMR-Werte wurde das Modell daher sowohl für die Eltern- wie auch Erzieherperspektive als plausibel beibehalten. Die in Tabelle 2 dargestellten standardisierten Faktorladungen der Items im entsprechenden spezifizierten Modell der Skala Expansives Verhalten an Replikationsstichprobe A liegen mit Werten von .70 bis .93 gleichmäßig hoch und können, da *Expansives Verhalten* als einziger latenter Faktor angenommen wurde, als Korrelation der Items mit der Skala interpretiert werden (vgl. Kline, 1998).

Anschließend wurde in der Kreuzvalidierung der Befunde zur einfaktoriellen Betrachtung der Skala *Expansi-*

ves Verhalten an der Replikationsstichprobe B geprüft, ob die Faktorladungen in dieser gegenüber denen in Replikationsstichprobe A invariant waren. Hierzu wurden zwei Varianten einer Mehrgruppenanalyse gerechnet und über einen χ^2 -Differenzentest miteinander verglichen. In der ersten Variante konnten die Faktorladungen in beiden Replikationsstichproben frei variieren, in der zweiten wurden die Faktorladungen der Replikationsstichprobe B denen der Replikationsstichprobe A gleichgesetzt. Nachdem sich in der konfirmatorischen Überprüfung der χ^2 -Differenzwert ($\Delta\chi^2$) der beiden verglichenen Varianten für beide Perspektiven auf dem 5 %-Niveau als nicht signifikant zeigte ($\Delta\chi^2_{ER} = 4.09, p > .05; \Delta\chi^2_{EL} = 12.65, p > .05$), wurde die Variante mit invarianten Faktorladungen im Sinne der „parsimony“-Regel bevorzugt.

Kennwerte der Items und Skalen in Köln und Braunschweig

Tabelle 3 zeigt Itemtrennschärfen und interne Konsistenzen der Skalen in den Gesamtstichproben aus Köln und Braunschweig. Insgesamt bestätigte sich der Eindruck aus den Untersuchungen an der Ausgangsstichprobe Köln I und den Voranalysen: Die Skala *Expansives Verhalten*

Tabelle 3. Interne Konsistenzen der Skalen des *PEP-Screen* und Trennschärfen der Items auf Subskalen und Gesamtskala (in Klammern) in den Gesamtstichproben aus Köln und Braunschweig

| Stichprobe | | Köln | | Braunschweig | |
|-------------|--|------------------------------|------------------------------|-----------------------------|-----------------------------|
| Perspektive | | ER | EL | ER | EL |
| EXPANSIV | 1. Streitet oder widerspricht viel | .67 (.60) | .47 (.47) | .63 (.55) | .45 (.43) |
| | 5. Kann sich nicht konzentrieren, kann nicht lange aufpassen | .59 (.57) | .45 (.46) | .52 (.46) | .48 (.51) |
| | 6. Ist unruhig überaktiv | .70 (.61) | .48 (.45) | .66 (.56) | .53 (.53) |
| | 8. Macht Sachen kaputt, die anderen gehören | .69 (.59) | .36 (.34) | .64 (.52) | .42 (.39) |
| | 10. Ist impulsiv oder handelt ohne zu überlegen | .78 (.65) | .54 (.48) | .76 (.65) | .59 (.53) |
| | 12. Greift andere körperlich an | .71 (.59) | .41 (.38) | .65 (.54) | .52 (.45) |
| | 13. Hat Wutausbrüche oder hitziges Temperament | .65 (.59) | .48 (.44) | .63 (.56) | .51 (.44) |
| | Interne Konsistenz | (N = 2830) $\alpha = .89$ | (N = 2133) $\alpha = .74$ | (N = 593) $\alpha = .86$ | (N = 520) $\alpha = .78$ |
| INTERNAL | 4. Klammert sich an Erwachsene oder ist zu abhängig | .43 (.26) | .35 (.33) | .41 (.23) | .37 (.35) |
| | 7. Ist zu furchtsam oder ängstlich | .48 (.11) | .38 (.23) | .51 (.12) | .38 (.18) |
| | 9. Ist unglücklich, traurig oder niedergeschlagen | .48 (.29) | .28 (.31) | .47 (.34) | .26 (.34) |
| | 11. Sagt häufig, dass ihm etwas weh tut (ohne körperliche Ursache) | .31 (.26) | .27 (.31) | .28 (.26) | .21 (.21) |
| | Interne Konsistenz | (N = 2830) $\alpha = .64$ | (N = 2136) $\alpha = .53$ | (N = 594) $\alpha = .63$ | (N = 521) $\alpha = .51$ |
| | 3. Verlangt viel Beachtung | – (.51) | – (.42) | – (.48) | – (.47) |
| | 2. Wird viel gehänselt | – (.38) | – (.25) | – (.29) | – (.19) |
| | Interne Konsistenz | (N = 2823) $\alpha = .82$ | (N = 2087) $\alpha = .75$ | (N = 518) $\alpha = .77$ | (N = 592) $\alpha = .79$ |

erwies sich im Erzieherurteil ($\alpha_K = .89/\alpha_{BS} = .86$) und im Elternurteil ($\alpha_K = .74/\alpha_{BS} = .78$) als zuverlässig. Die Trennschärfen der Items lagen im Erzieherurteil in Köln ($r_{itK} = .59 - r_{itK} = .78$) und in Braunschweig ($r_{itB} = .52 - r_{itB} = .76$) im stabilen bis guten Bereich. Ähnlich, wenn auch etwas niedriger, fielen die Kennwerte im Elternurteil in Köln ($r_{itK} = .36 - r_{itK} = .54$) und in Braunschweig ($r_{itB} = .42 - r_{itB} = .59$) aus. Im Erzieherurteil führte eine Skalenbildung aus den verbliebenen vier Items zu *Internalisierendem Verhalten* zu Zuverlässigkeitskennwerten, die nur knapp die Skalenbildung für gruppenstatistische Analysen rechtfertigen ($\alpha_K = .64$ bzw. $\alpha_{BS} = .63$), für das Elternurteil war dies nicht mehr gegeben ($\alpha_K = .53$ bzw. $\alpha_{BS} = .51$).

Hinweise auf Stabilität und Validität der Erfassung expansiven Verhaltens

Kinder, deren Wert auf der Summenskala *Expansives Verhalten* aus beiden Beurteilerperspektiven einen Prozentrang von über 85 erreichte, wurden in Orientierung an den Prävalenzraten (5 %–15 %) und dem von Achenbach für die übergeordneten Skalen seiner Instrumente empfoh-

lenen „borderline range“ von $\geq PR83$ (Achenbach & Rescorla, 2000) als Zielgruppe für die Evaluation des PEP definiert (Hanisch et al., 2006). Der anhand der Verteilung der Analysestichprobe Köln I festgestellte Wert von ≥ 14 Punkten indizierte auf der Basis der Gesamtstichprobe in Köln, für die beide Beurteilerperspektiven vorlagen ($N = 2123$) 260 (12,3 %) der Kinder, während bei Anwendung eines für Eltern bzw. Erzieher spezifischen Rohwert-Cut-Offs (≥ 6 bzw. ≥ 7 Punkte) beim gleichen Prozentrang aus beiden Perspektiven übereinstimmend („Schnittmengenmodell“) nur 40 % dieser Kinder indiziert worden wären. Damit stellt das gewählte Vorgehen ein im Sinne des präventiven Ansatzes günstigeres, weil weniger strenges Kriterium dar.

Die Positive Prädiktive Kraft (PPK) erfasst den Anteil der tatsächlich „Auffälligen“ an den als solchen indizierten Kindern. Aus Fragebogendaten der Eltern und Erzieherinnen sowie den Angaben in den Diagnosechecklisten der Untersucher beim Hausbesuch (DISYPS-KJ, Döpfner & Lehmkuhl, 2000), wurde die Erfüllung der Diagnosekriterien nach DSM-IV überprüft. Dabei mussten einem Kind nach mindestens einer der drei Perspektiven (Eltern, Erzieher, klinische Beurteiler) eine oppositionelle Störung und

Tabelle 4. Korrelationen der Skala *Expansives Verhalten* aus dem PEP-Screen mit inhaltlich konvergenten bzw. divergenten Skalen aus dem Prätest (indizierte Gruppe Köln; $N = 155$)

| | | PEP-Screen: Expansives Verhalten | | |
|--|--------------|----------------------------------|----------------------------|--------------------------|
| Prätest-Skala | | Erzieherin | Eltern | Gesamt |
| rekonstruierte Screening-Skala aus CBCL/C-TRF <i>Expansives Verhalten</i> ¹ | Erzieherin | .49*** | -.09 | .39*** |
| | Mutter | -.06 | .46*** | .34*** |
| | Gesamt | .31*** | .25** | .51*** |
| C-TRF bzw. CBCL-Syndromskala <i>Externalisierendes Verhalten</i> | Erzieherin | .46*** | -.07 | .39*** |
| | Mutter | -.01 | .47*** | .31*** |
| C-TRF bzw. CBCL-DSM Skala <i>ADHD</i> | Erzieherin | .41*** | -.05 | .35*** |
| | Mutter | -.03 | .42*** | .33*** |
| C-TRF bzw. CBCL-DSM-Skala <i>ODD</i> | Erzieherin | .32*** | -.03 | .28 |
| | Mutter | -.06 | .38*** | .28*** |
| Fremdbeurteilungsbogen HKS <i>HKS-Gesamtskala</i> | Erzieherin | .37*** | -.10 | .27*** |
| | Mutter | -.05 | .40*** | .30*** |
| | Diagnostiker | -.04 | .18* | .12 |
| Fremdbeurteilungsbogen SSV <i>Oppositionelles Verhalten</i> | Erzieherin | .45*** | .04 | .46 |
| | Mutter | -.10 | .45*** | .29*** |
| | Diagnostiker | -.06 | .25** | .16 |
| C-TRF- bzw. CBCL-Syndromskala <i>Internalisierendes Verhalten</i> | Erzieherin | .08 (–.18*) ^a | .05 (.10) ^a | .12 (–.07) ^c |
| | Mutter | –.06 (.01) ^b | .31*** (–.02) ^b | .21** (.04) ^c |

Anmerkungen: * $p \leq .05$; ** $p \leq .01$; *** $p \leq .001$. ¹ Statt Item 1 „Streitet oder widerspricht viel“ C-TRF20 „Gehorcht nicht“, Statt Item 10 „Ist impulsiv oder handelt ohne zu überlegen“ C-TRF59 „Wechselt sehr schnell von einer Beschäftigung zur nächsten“. ^a Partialkorrelation nach herausrechnen der Skala Externalisierendes Verhalten (Erzieher; Prätest). ^b Partialkorrelation nach herausrechnen der Skala Externalisierendes Verhalten (Mutter; Prätest). ^c Nach herausrechnen der Skala Externalisierendes Verhalten (Erzieherin und Mutter; Prätest).

aus mindestens zwei der drei Perspektiven eine Aufmerksamkeitsdefizit-/Hyperaktivitätsstörung (ADHS) zugeschrieben werden und auch die erfassten Zusatzkriterien erfüllt sein. 31 der indizierten und erreichten 155 Kinder (20 %) erhielten auf diese Weise eine Diagnose. Ein weiches kategoriales Kriterium stellt die Anwendung der von Achenbach empfohlenen Prozentranggrenzen auf die übergeordneten Skalen (*Internalisierendes* und *Externalisierendes Verhalten*, *Gesamtwert*) in der CBCL/1½–5 bzw. in der C-TRF/1½–5 dar, die für den deutschen Sprachraum anhand von umfangreichen repräsentativen Stichproben (Plück et al., in Vorbereitung) vorläufig bestimmt werden konnten. Auf Basis des PR 90 („clinical range“), ergaben sich nach Elternurteil 29 %, nach Erzieherurteil sogar 44.5 % Kinder mit auffällig externalisierendem Verhalten. Die Raten für auffälliges internalisierendes Verhalten fielen mit 14.8 % bzw. 18.1 % hingegen wie erwartet geringer aus.

Eine Überprüfung der Mittelwertunterschiede (*t*-Test) zwischen den Kindern mit bzw. ohne Diagnose hinsichtlich der Screening-Skala *Expansives Verhalten* aus Sicht der Eltern ($M(SD) = 8.94(2.13)$ bzw. $M(SD) = 7.47(2.49)$), der Erzieherinnen ($M(SD) = 10.48(3.04)$ bzw. $M(SD) = 9.35(2.65)$) bzw. der übergreifenden Skala ($M(SD) = 19.42(3.41)$ bzw. $M(SD) = 16.81(2.51)$) zeigte Abweichungen in der erwarteten Richtung zwischen beiden Gruppen auf. Die Unterschiede auf der Gesamtskala ($t = -3.987; p = .000$) und in der Elternperspektive ($t = 3.016; p = .003$) fielen hoch signifikant aus und entsprachen mit $d = .68$ bzw. $d = .63$ einem starken Effekt (Cohen, 1988). Hinsichtlich der Erzieherperspektive war dieser Unterschied zwischen indizierten Kindern mit und ohne Diagnose nur gering ($d = .36$) ausgeprägt und wurde auch nur auf dem 5 %-Niveau statistisch signifikant ($t = -2.073; p = .040$).

Korrelationen mit dimensional Maßen des kindlichen Verhaltens im „Prätest“, die eher ähnliches Verhalten erfassen, können als Maß für die Stabilität der Beurteilung durch Eltern und Erzieherin (Zeitabstand etwa 2 bis 3 Monate) sowie für Konvergenz betrachtet werden, während solche mit Dimensionen eher anderen Verhaltens auf Divergenzen verweisen. In Tabelle 4 sind die ermittelten Zusammenhangsmaße aufgeführt.

Erwartungsgemäß fielen die Korrelationen der Screening-Skala *Expansives Verhalten* mit Skalen, die ebenfalls verschiedene Aspekte expansiven Verhaltens erfassen sicher signifikant aus und lagen höher als die Korrelationen mit der Skala *Internalisierendes Verhalten*, insbesondere dann, wenn sie in der gleichen Beurteilerperspektive betrachtet wurden. Die maximale Korrelation von $r = .51$ wurde zwischen der zur Indizierung eingesetzten Gesamtskala *Expansives Verhalten* und der „rekonstruierten“ korrespondierenden Skala aus dem Prätest erreicht (s. Instrumente), was als Maß für die Retest-Reliabilität herangezogen werden kann.

Die Korrelationen mit der Skala *Internalisierendes Verhalten* fielen in beiden Beurteilerperspektiven wie erwartet zwar niedriger aus, jedoch bestanden in der Elternper-

spektive zwischen dieser und der Screening-Skala *Expansives Verhalten* ($r = .31$) sowie mit der Informanten übergreifenden Sicht ($r = .21$) positive statistisch signifikante Korrelationen. Diese ließen sich jedoch fast vollständig durch die Zusammenhänge zur Skala *Externalisierendes Verhalten* aus dem „Prätest“ erklären, wie die dargestellten Partialkorrelationen ($r_p = .02$ bzw. $r_p = .04$) zeigen.

Diskussion

Die Bewertung des *PEP-Screen* als Instrument zur Früherkennung expansiven Verhaltens soll anhand des Teils der entsprechenden Kriterien von Wilson und Jungner (1968) erfolgen, die speziell an solche Untersuchungsverfahren angelegt werden:

Eignung: Die faktorielle Validität der Skala *Expansives Verhalten* konnte an verschiedenen Stichproben trotz bestehender Unterschiede in relevanten soziodemographischen Variablen (Sozialer Status, soziale Belastung) repliziert werden. Dies galt auch für die internen Konsistenzen der Skala, so dass dieser Kern des Instruments daher als bestätigt betrachtet werden kann. Darüber hinaus erwies sich auch die Gesamtskala Stichproben übergreifend als konsistent. Dies spricht insgesamt für die Robustheit des Instruments. Diese Einschätzung wird auch durch die konfirmatorischen Analysen am 2-faktoriellen Modell gestützt, wenngleich diese und die Befunde zu den internen Konsistenzen die Verwendung der Items zu internalisierenden Verhaltensweisen als eigene Skala nur im Erzieherurteil für gruppenstatistische Analysen bestätigten. Für die Skala *Expansives Verhalten* wurden mit Blick auf die Beurteiler übergreifenden Zusammenhänge sowohl in der Kölner als auch in der Braunschweiger Gesamtstichprobe nur minimale Korrelationen zwischen beiden Perspektiven von $r = .38$ festgestellt. Dieses Ergebnis steht in Korrespondenz zu den berichteten Befunden anderer Studien und bestätigt die Notwendigkeit beide Perspektiven bei der Indizierung zu berücksichtigen. Dies gilt um so mehr, als sowohl Erzieherinnen als auch Eltern Ziel der angebotenen indizierten Präventionsmaßnahme (PEP) waren. Die Eignung des Instruments zur Erfassung expansiven Verhaltens wird durch die gefundenen Korrelationen zu verschiedenen konvergenten Konstrukten unterstrichen, während sich die Zusammenhänge zum divergenten Konstrukt *Internalisierendes Verhalten* unter Berücksichtigung der Partialkorrelationen minimieren ließen. Hinweise auf die Stabilität dieser Messung werden durch die bedeutsamen Korrelationen zu der zeitlich später erfassten „rekonstruierten“ Skala belegt. Hinweise auf die Eignung der Beurteiler übergreifenden Skala *Expansives Verhalten* zur Indizierung anhand des gewählten Cut-Offs ergeben sich aus der festgestellten prädiktiven Kraft. Diese fiel in Abhängigkeit vom gewählten Kriterium unterschiedlich deutlich aus. Natürlich hat der gewählte Cut-Off Einfluss auf die Höhe der Übereinstimmung zwischen Indizierung und Diagnose. Ziel war aber eben nicht die Früherkennung von späteren Diagnosen, sondern die Auswahl einer Stichprobe mit erhöhtem Risiko. Nach verschiede-

nen Studien aus der in dem Bereich führenden Conduct Problem Prevention Research Group (Lochmann & CPPRG, 1995; Hill et al., 2004) kann auch eine Steigerungsrate der Versorgung von 20 % (klinische Inanspruchnahme) auf über 40 % durch Indizierung schon als bedeutender Fortschritt betrachtet werden. Jedenfalls wies die Dimension *Expansives Verhalten* in der genauer untersuchten Gruppe zwischen Kindern mit und ohne Diagnose signifikante Unterschiede auf und differenziert also auch oberhalb des gewählten Cut-Offs insbesondere unter Berücksichtigung des Elternurteils.

Aussagen zur *Akzeptanz in der Bevölkerung* ergeben sich aus der – wenn auch über die Stichproben differierenden – Rate der Eltern, die sich an der Erhebung des Screenings beteiligten. Im ungünstigsten Falle waren immerhin drei Viertel der Eltern bereit, die über die Erzieherinnen der Kinder vom Kindergarten ausgehend verteilten Fragebögen auszufüllen; dies allerdings unter der Bedingung, dass freigestellt war, vollständig anonym zu bleiben, und damit auch keine Rückmeldung über ein Ergebnis zu erhalten bzw. eine evtl. angebotene Maßnahme in Anspruch zu nehmen. Allerdings wurden nur 6,5 % der indizierten Kinder durch diese Bedingung aus dem weiteren Studienverlauf ausgeschlossen. Die Mehrheit der anderen Ablehner gab trotz Erreichbarkeit vielfältige andere Gründe an; am häufigsten müssen aus den unvollständigen Angaben Fremdsprachigkeit, und Datenschutzgründe angenommen werden. Außerdem handelte es sich um solche Kinder, die von Ihren Eltern als eher weniger auffällig, von ihren Erzieherinnen als eher auffälliger beurteilt worden waren, als die, die in der Studie verblieben.

Die Kürze und Praktikabilität des Fragebogens ließe die *Möglichkeit kontinuierlicher Untersuchungen* zu. Im Rahmen der vorliegenden Studie wurden querschnittlich alle Kinder einer jeden Gruppe erfasst. Denkbar wäre eine systematische Erfassung zu Stichtagen im Jahr oder nach Jahrgängen. Bewährt hat sich das Instrument jedenfalls beim Einsatz in überschaubaren Stichproben, in denen dann dem indizierten Anteil auch ein entsprechendes präventives Angebot gemacht werden konnte.

Grenzen

Nicht berücksichtigt wurde in der vorliegenden Studie die alters- und insbesondere geschlechtsspezifische Verteilung des Störungsbildes. Auf diese Weise enthielt die indizierte Gruppe bei der bestehenden „Knabenwendigkeit“ automatisch mehr Jungen als Mädchen, was aber im Sinne des präventiven Vorgehens sinnvoll war.

Um einen Screening-Fragebogen hinsichtlich externalisierenden Verhaltens letztendlich wirksam prüfen zu können, wäre zusätzlich eine Stichprobe nötig, die die gesamte Bandbreite der möglichen Ausprägungen repräsentiert. Das war in der vorliegenden Studie für die Ausgangsstichprobe auch der Fall. Aus projektökonomischen Gründen wurde jedoch darauf verzichtet, einen angemessenen großen und ebenfalls repräsentativen Teil dieser Stich-

probe zum gleichen Zeitpunkt wie die indizierte Gruppe weiter zu verfolgen, da der Schwerpunkt der Datenerhebung der Evaluation des indizierten Präventionsprogramms für Expansives Problemverhalten (PEP) galt. Die vorliegende Arbeit ist daher methodisch bedingt in ihrer Aussagekraft hinsichtlich der Eignung des Fragebogens zur Indizierung eingeschränkt. Effekte des Alters, der Wahl des Cut-Offs oder des Erhebungszeitpunktes im Jahr blieben ebenfalls außen vor. Darüber hinaus betreffen die Aussagen zur Gruppe der indizierten Kinder nur diejenigen $N = 155$ Fälle von den ursprünglich $N = 260$, die vollständig untersucht werden konnten (ca. 60 %). Unter Anlegen der strengen Diagnosekriterien ergaben sich damit ca. 2,5 % Kinder mit einer klinisch relevanten Störung in der Gesamtstichprobe. In vielen epidemiologischen Studien liegen die Raten, allerdings auch in Abhängigkeit von der Strenge der Kriterien, höher. Es ist daher davon auszugehen, dass mit dem gewählten Screening-Instrument und -Kriterium „Falsch Negative“ Indizierungen entstanden sind, die im Sinne des präventiven Ansatzes von besonderer Bedeutung sind. Wie sich z. B. bei Breuer und Döpfner (2006) in einer deutschen Feldstichprobe dieser Altersgruppe zeigte, liegt die Übereinstimmung der Diagnosevergabe anhand der Symptomkriterien zwischen Eltern und Erziehern vergleichsweise niedrig (4,2 % gegenüber 11,3 % bzw. 6,6 % auf). In der vorliegenden Studie wurde das Kriterium „Situationsübergreifende Ausprägung“ für die Diagnose ADHS sehr streng im Sinne einer Erfüllung der vollständigen Symptomkriterien in mindestens 2 Perspektiven ausgelegt und schloss neben den reinen Symptomkriterien auch die erfassten Zusatzkriterien ein.

Ausblick

Zur Beantwortung offener Fragen wären weitere umfangreichere Erhebungen gerade auch an einem repräsentativen Anteil der im Screening unauffälligen Stichprobe insbesondere für ROC-Analysen zur optimierten Festlegung des Cut-Offs nötig, um den Einsatz dieses Instruments für ein „Bevölkerungs-Screening“ zu rechtfertigen. In dem oben beschriebenen Einsatz und für die kompakte dimensionale Erfassung expansiven Verhaltensweisen im Vorschulalter hat sich der *PEP-Screen* jedoch bewährt.

Literatur

- Achenbach, T. M. & Rescorla, L. (2000). *Manual for the ASEBA preschool forms and profiles*. Burlington, VT: University of Vermont Department of Psychiatry.
- Arbeitsgruppe Deutsche Child Behavior Checklist (1998). *Elternfragebogen über das Verhalten von Kindern und Jugendlichen; deutsche Bearbeitung der Child Behavior Checklist (CBCL/4-18). Einführung und Anleitung zur Handauswertung* (2. Aufl. mit deutschen Normen, bearbeitet von M. Döpfner, J. Plück, S. Bölte, K. Lenz, P. Melchers & K. Heim). Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).
- Arbeitsgruppe Deutsche Child Behavior Checklist (2000a). *Elternfragebogen für Klein- und Vorschulkinder (CBCL/*

- 1½-5). Köln: Arbeitsgruppe Kinder-, Jugend- und Familien-diagnostik (KJFD).
- Arbeitsgruppe Deutsche Child Behavior Checklist (2000b). *Fragebogen für ErzieherInnen von Klein- und Vorschulkindern (C-TRF/1½-5)*. Köln: Arbeitsgruppe Kinder-, Jugend- und Familiendiagnostik (KJFD).
- Bennett, K. J., Lipman, E. L., Brown, S., Racine, Y., Boyle, M. H. & Offord, D. R. (1999) Predicting Conduct Problems: Can high-risk children be identified in kindergarten and grade 1? *Journal of Consulting and Clinical Psychology*, 67 (4), 470–480.
- Bennett, K. J., Lipman, E. L., Racine, Y. & Offord, D. R. (1998) Do measures of externalizing behavior in normal populations predict later outcomes? Implications for targeted interventions to prevent conduct disorder. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 39, 1059–1070.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Breuer, D. & Döpfner, M. (2006). Aufmerksamkeitsdefizit-/Hyperaktivitätsstörungen bei Drei- bis Sechsjährigen in der ärztlichen Praxis – eine bundesweite Befragung. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie*, 34 (5), 357–365.
- Cai, X., Kaise, A. P. & Hancock, T. B. (2004) Parent and teacher agreement on child behavior checklist items in a sample of preschoolers from low-income and predominantly African American families. *Journal of Clinical Child and Adolescent Psychology*, 33 (2), 303–312.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). New York: Academic Press.
- Döpfner, M. & Lehmkuhl, G. (1997). Von der kategorialen zur dimensional Diagnostik. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 46, 519–547.
- Döpfner, M. & Lehmkuhl, G. (2000). *Diagnostik-System für Psychische Störungen im Kindes- und Jugendalter nach ICD-10 und DSM-IV (DISYPS-KJ)*. (2. Aufl.). Bern: Huber.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581–586.
- Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. (2003). Using the strengths and difficulties questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *International Review of Psychiatry*, 15, 166–172.
- Hanisch, C., Plück, J., Meyer, N., Brix, G., Freund-Braier, I. & Hautmann, C., et al. (2006). Kurzeffekte des indizierten Präventionsprogramms für Expansives Problemverhalten (PEP) auf das elterliche Erziehungsverhalten und auf das kindliche Problemverhalten. *Zeitschrift für Klinische Psychologie und Psychotherapie*, 35 (2), 117–126.
- Hill, L. G., Lochmann, J. E., Coie, J. D. & Greenberg, M. T. (2004). Effectiveness of early screening for externalizing problems: Issues of screening accuracy and Utility. *Journal of Consulting and Clinical Psychology*, 72 (5), 809–820.
- Hu, L. & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (S. 76–99). Thousand Oaks, CA: Sage.
- Hudziak, J. J., Copeland, W., Stanger, C. & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a Receiver-Operating Characteristic Analysis. *Journal of Child Psychology and Psychiatry*, 45 (7), 1299–1307.
- Ihle, W., Esser, G., Laucht, M. & Schmidt, M. H. (2004). Depressive Störungen und aggressiv-dissoziale Störungen im Kindes- und Jugendalter – Prävalenz, Verlauf und Risikofaktoren. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz*, 47, 728–735.
- Jöreskog, K. G. & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Chicago, ILL: Scientific Software.
- Kasius, M. C., Ferdinand, R. F., van den Berg, H. & Verhulst, F. C. (1997). Associations between different diagnostic approaches for child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 38, 1591–1594.
- Kline, R. B. (1998). *Principles and practice of structural equation modelling*. New York: Guilford Press.
- Lampert, T. L., Polanczyk, G., Tramontina, S., Mardini, V. & Rohde, L. A. (2004). Diagnostic performance of the CBCL-Attention Problem Scale as a screening measure in a sample of Brazilian children with ADHD. *Journal of Attention Disorders*, 8 (2), 63–71.
- Lochman, J. E. & CPPRG (1995). Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology*, 63 (4), 549–559.
- Loeber, R. (1990). Development and risk factors of juvenile antisocial behavior and delinquency. *Clinical Psychology Review*, 10, 1–41.
- Maniadaki, K., Karaba, R. & Kakouros, E. (2005). The use of the strengths and difficulties questionnaire (SDQ) as a screening instrument for behaviour problems in preschoolers. *Poster at the XXXV Annual Congress of the EABCT, Thessaloniki Greece*.
- Miller, Y., Kuschel, A. & Hahlweg, K. (2002). Frühprävention von externalisierenden Störungen – Grundprinzipien und elternzentrierte Ansätze zur Prävention von expansiven kindlichen Verhaltensstörungen. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 51, 411–453.
- Muir Gray, J. A. (1994). Testing a Test. *Bandolier – Evidence based thinking about health care*, 3, 1–2.
- Mullick, M. S. & Goodman, R. (2001). Questionnaire screening for mental health problems in Bangladeshi children: A preliminary study. *Social Psychiatry and Psychiatric Epidemiology*, 36, 94–99.
- Plück, J., Döpfner, M., Kuschel, A., Heinrichs, N., Denner, C. & Schmeck, K. (in Vorbereitung). *Fragebogen für Eltern und Fragebogen für Erzieherinnen von Klein- und Vorschulkindern (CBCL 1½-5/C-TRF 1½-5) – Normwerte auf der Basis deutscher Stichproben*. Unveröffentlichtes Manuskript, Universität zu Köln.
- Plück, J., Wiczorrek, E., Wolff Metternich, T. & Döpfner, M. (2006). *Präventionsprogramm für Expansives Problemverhalten (PEP). Ein Manual für Eltern- und Erziehergruppen*. Göttingen: Hogrefe.
- Reid, J. R. (1993). Prevention of conduct disorder before and after school entry: Relating intervention to developmental findings. *Development and Psychopathology*, 5, 243–262.
- Rescorla, L. A. (2005) Assessment of young children using the Achenbach system of empirically based assessment (ASE-BA). *Mental Retardation and Developmental Disabilities Research Reviews*, 11, 226–237.
- Van Lier, P. A., Verhulst, F. C., van der Ende, J. & Crijnen, A. A. (2003). Classes of disruptive behaviour in a sample of young elementary school children. *Journal of Child Psychology and Psychiatry*, 44 (3), 377–387.
- Verhulst, F. C. & Akkerhuis, G. W. (1989). Agreement between parents' and teachers' ratings of behavioral/emotional problems of children aged 4–12. *Journal of Child Psychology and Psychiatry*, 30, 123–136.
- Werthamer-Larsson, L., Kellam, S. G. & Wheeler, L. (1991). Effects of first-grade-classroom environment on shy behaviour, aggressive behavior, and concentration problems. *American Journal of Community Psychology*, 19, 585–602.
- Wilson, J.M.G. & Jungner, F. (1968). *Principles of screening for disease*. Geneva: World Health Organisation, Public Health Papers, No. 34.

Dr. Julia Plück, Dipl.-Psych., Dr. Christopher Hautmann
Gabriele Brix, Inez Freund-Braier
und Prof. Dr. Manfred Döpfner

Klinik und Poliklinik für Psychiatrie und Psychotherapie
des Kindes- und Jugendalters der Universität zu Köln
Robert-Koch-Straße 10
50931 Köln

Prof. Dr. Kurt Hahlweg

Institut für Psychologie
der TU Braunschweig
Postfach 33 29
38023 Braunschweig