

**Standardisierte Persönlichkeitsdiagnostik und Behandlungsevaluation:
Methodenkritische Analyse der Handhabung im Strafvollzug**

Von der Philosophischen Fakultät
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des Grades eines
Doktors der Philosophie (Dr. phil.) genehmigte
Dissertation

von Oliver Lauterbach,
geboren am 21.10.1971 in Hannover

2011

Referent: Prof. Dr. Alois Wacker
Korreferentin: Prof. Dr. Adelheid Kühne
Tag der mündlichen Prüfung: 10.01.2011

Danksagung

Ich danke Prof. Dr. Daniela Hossler für die Möglichkeit, diese Arbeit im Projekt „Entwicklungsfolgen der Jugendstrafe“ anfertigen zu können sowie ihre Beratung und stetige Ermutigung. Meine Kolleginnen Verena Boxberg und Anabel Taefi haben mir in zahllosen Diskussionen wichtige Hinweise gegeben. Alle Mitarbeiter des Kriminologischen Forschungsinstituts Niedersachsen boten mir über die Jahre ein äußerst hilfreiches und auch in schwierigen Zeiten sehr kollegiales Arbeitsumfeld. Ganz besonders möchte ich Dr. Christiane Bosold danken, ohne deren Unterstützung insbesondere in der Endphase ich die Arbeit vermutlich nicht fertig gestellt hätte. Nicht zuletzt danke ich meinem Doktorvater Prof. Dr. Alois Wacker für seine Betreuung und Prof. Dr. Adelheid Kühne für die Erstellung des Zweitgutachtens.

Zusammenfassung

In der vorliegenden Arbeit werden methodische Schwierigkeiten der Behandlungsevaluation im Jugendstrafvollzug anhand verschiedener empirischer Untersuchungen beleuchtet. Als ein Themenbereich wird die Verwendung standardisierter Instrumente der Persönlichkeitsdiagnostik in Behandlungsuntersuchung und Behandlungsevaluation erörtert. Befunde zeigen, dass das Freiburger Persönlichkeitsinventar sowie der Interpersonal Reactivity Index zur Erfassung von Empathie nur eingeschränkt mit jungen Inhaftierten eingesetzt werden können. Ein zweiter Themenbereich umfasst die methodische Güte von Untersuchungsdesigns der Behandlungsforschung. Anhand mehrerer Untersuchungen von Behandlungsmaßnahmen für Gewalt- und Sexualstraftäter wird aufgezeigt, dass Evaluationen trotz der schwierigen Absicherbarkeit insbesondere ihrer internen Validität praxisrelevante Schlussfolgerungen erlauben. Als drittes Themenfeld wird die Verwendung von Erfolgskriterien der strafvollzuglichen Behandlung diskutiert. Hier wird argumentiert, dass neben dem Rückfallkriterium und der Verwendung interventionsnaher Indikatoren zusätzlich solche der sozialen Einbindung nach Haftentlassung verwendet werden sollten.

Schlagwörter: Jugendstrafvollzug, Persönlichkeitsdiagnostik, Behandlungsevaluation

Abstract

In this paper, methodological difficulties of treatment evaluation in the German youth prison system are discussed referring to several empirical studies. As a first topic, the usability of standardized personality measures used for treatment planning and treatment evaluation is analyzed. There is evidence that broad-band personality measures (like the Freiburger Persönlichkeitsinventar) and specific measures (like the Interpersonal Reactivity Index as an empathy scale) may be of limited validity when used with young prison inmates. As a second topic, methodological quality of experimental design in treatment evaluation is discussed. Referring to studies with young violent and sexual offenders it is argued, that although adequate internal validity can often not be assured, findings and conclusions of practical relevance can be obtained. As a third topic, effect-measures of offender treatment are considered. The argument is that in addition to reoffense-rates and measures taken directly after treatment, the degree of social inclusion after release from prison should be used to measure treatment effectiveness.

Keywords: Youth prison system, Personality assessment, Treatment evaluation

Inhalt

1	Einleitung	7
2	Hintergrund	9
2.1	Evaluation in der Kriminalprävention.....	10
2.2	Kenntnisstand der Forschung zur Straftäterbehandlung	12
2.3	Behandlung im Jugendstrafvollzug	14
3	Eigene empirische Untersuchungen	16
3.1	Einsatz standardisierter Persönlichkeitstests in der Straftäterdiagnostik	18
3.1.1	Einsatz des FPI-R bei Inhaftierten	21
3.1.2	Erfassung von Empathie bei Inhaftierten	22
3.2	Güte von Untersuchungsdesigns in der Behandlungsforschung	23
3.2.1	Psychosoziale Behandlung im Jugendstrafvollzug	29
3.2.2	Sozialtherapeutische Behandlung von jungen Sexualstraftätern	31
3.2.3	Evaluation eines Anti-Aggressivitäts-Trainings	32
3.3	Effektkriterien der Behandlungsforschung	34
3.3.1	Entlassungsvorbereitung und Nachentlassungssituation.....	38
3.3.2	Soziale Integration und Delinquenz nach Haftentlassung.....	39
4	Fazit und Ausblick	40
4.1	Perspektiven für den Jugendstrafvollzug	40
4.2	Perspektiven für die Behandlungsforschung im Strafvollzug.....	42
	Literatur	45
	Teile der kumulativen Dissertation	61

1 Einleitung

Psychologische Diagnostik sowie Evaluationsforschung sind für den modernen Strafvollzug unverzichtbar, werden in diesem Anwendungsbereich jedoch vor besondere Herausforderungen gestellt. Gängige Selbstbeurteilungsverfahren zur Erhebung von Persönlichkeitsmerkmalen stellen hohe Anforderungen an kognitive und sprachliche Fähigkeiten der Untersuchten und bergen in einem Zwangskontext die Gefahr von Reaktivität, dessen ungeachtet werden sie in der Behandlungsuntersuchung standardmäßig zu Zwecken der Vollzugsplanung eingesetzt. Auch die Evaluation strafvollzuglicher Behandlungsmaßnahmen orientiert sich nicht in erster Linie an fachlichen Gütestandards, sondern wird in den meisten Fällen pragmatisch an Vollzugsgegebenheiten angepasst, woraus sich Konsequenzen für die Validität der Ergebnisse ergeben. Weiterhin ist die Aussagekraft von Evaluationsstudien an die Art der verwendeten Erfolgskriterien gekoppelt, die entweder in proximalen Maßen wie Veränderung personbezogener Merkmale oder distalen Maßen wie Rückfälligkeit bestehen und jeweils spezifische Vor- und Nachteile mit sich bringen. In der vorliegenden Arbeit werden diese Probleme anhand empirischer Untersuchungen näher beleuchtet und diskutiert.

Nach den gesetzlichen Vorgaben dient die Behandlungsuntersuchung im Strafvollzug der Erforschung der Lebensverhältnisse und der Persönlichkeit des Gefangenen zum Zweck der Erstellung eines Vollzugsplans. Dies geschieht nicht allein nach Aktenlage, sondern die originäre Erkenntnisquelle ist der Gefangene selbst, der unter Verwendung wissenschaftlich abgesicherter Methoden zu untersuchen ist (Laubenthal, 2003). Hierzu werden unter anderem etablierte und normierte psychodiagnostische Instrumente wie z.B. Persönlichkeitsinventare und Intelligenztests eingesetzt. Während sich bei der Intelligenzdiagnostik eine weitgehende Kulturunabhängigkeit (u.a. Freiheit von sozialem Status und Milieu) realisieren lässt, ist dieser Anspruch bei Persönlichkeitsinventaren schwerer zu erfüllen. Hier muss die Frage gestellt werden, ob für die Allgemeinpopulation entwickelte Testverfahren an einer so spezifischen Bevölkerungsgruppe wie Straftätern, die sich beispielsweise durch unterdurchschnittliches Bildungsniveau, eine Kumulation negativer Sozialisationserfahrungen und möglicherweise geringer ausgeprägte Introspektions- und Reflexionsfähigkeit auszeichnen, ausreichend valide sind. Eine andere Frage ist, ob neben der grundsätzlichen Eignung der Instrumente die Erhebungssituation die Verwertbarkeit der Ergebnisse beeinflusst. Gerade aufgrund der Relevanz der Untersuchungsergebnisse für den weiteren Vollzugsverlauf muss mit hoher Tendenz zu sozial erwünschtem Antwortverhalten bis hin zu gezielter Verfälschung gerechnet werden. Die Bedeutung der Eingangsdagnostik im Strafvollzug lässt sich aus Ergebnissen von Meta-

analysen ableiten, die bei Beachtung des sogenannten Risiko-, Bedürfnis- und Ansprechbarkeitsprinzips eine besonders hohe Behandlungseffektivität belegen (Andrews, Bonta & Hoge, 1990). Die Passung von Intensität, Ausrichtung auf individuelle Defizite sowie die Berücksichtigung von Lernfähigkeit und Motivation der Inhaftierten setzen allerdings die genaue Diagnose eben dieser Merkmale voraus und begründen somit die Bedeutung einer ausreichend reliablen und validen Eingangsdagnostik für die Zuweisung zu erfolgversprechenden Behandlungsmaßnahmen.

Neben der Sicherstellung einer möglichst optimalen Zuweisung zu Behandlungsmaßnahmen im Strafvollzug ist für die Maßnahmenplanung und Weiterentwicklung die Evaluation der Behandlungseffektivität von großer Bedeutung. In Deutschland wurde in den 80er Jahren eine Debatte um die Evaluation der damals neu geschaffenen sozialtherapeutischen Anstalten geführt (vgl. Kury, 1983a), die bis heute die intensivst mögliche Form von Behandlung im Strafvollzug darstellen. Probleme der Evaluation solcher Institutionen wie auch einzelner Behandlungsmaßnahmen im normalen Vollzug bestehen unter anderem in der Kontrollgruppenproblematik und dem Wirksamkeitskriterium. Die Kontrollgruppenproblematik ergibt sich aus der im Strafvollzug schwierigen Realisierbarkeit experimenteller Untersuchungsdesigns und dem Vorherrschen von Untersuchungsdesigns geringerer Güte. Dies liegt unter anderem daran, dass die bereits erwähnte Behandlungszuweisung notwendigerweise in Abhängigkeit von individuellen Merkmalen der Inhaftierten erfolgt und nicht, wie es aus Sicht des Evaluationsforschers wünschenswert wäre, randomisiert. Ein Beispiel hierfür ist die Behandlung von Sexualstraftätern, die seit der Gesetzesänderung ab 2003 bei vorliegender Indikation unbedingt in eine sozialtherapeutische Anstalt bzw. Abteilung zu verlegen sind (Dünkel, 2005). Eine Kontrollgruppe Unbehandelter lässt sich so nicht realisieren und Abbrecher oder für die Behandlung nicht geeignete Gefangene weisen als Vergleichsgruppen systematische Unterschiede zu den Behandelten auf. In der Folge stellen die in Behandlungsevaluationen akquirierten Kontrollprobanden in vielen Fällen keine adäquate Vergleichsgruppe zu den Behandelten dar.

Das Wirksamkeitskriterium der Straftäterbehandlung ist üblicherweise die Legalbewährung, also das Ausbleiben (meist offiziell registrierter) erneuter Delinquenz. Neben diesem Kriterium werden auch Merkmale der sozialen Integration nach Haftentlassung (die sogenannte Sozialbewährung; Ortmann, 2002) und Veränderungen psychologischer kriminalitätsrelevanter Merkmale als zwischengelagerte Wirkfaktoren angenommen. Die Verwendung unterschiedlicher Effektivitätskriterien bedeutet vor dem Hintergrund der Evaluationsforschung die Unterscheidung von Ansätzen, die allein an der Zielerreichung orientiert sind („black-box evaluation“) von solchen, die auch Prozessmerkmale berücksichtigen und von

einem Wirkmodell ausgehen („theory-driven evaluation“; Chen & Rossi, 1983). Werden neben der offiziell registrierten Rückfälligkeit alternative Effektivitätskriterien von Strafvollzug oder strafvollzuglicher Behandlung herangezogen, so stellt sich allerdings die Frage, welche Bedeutung diesen im Vergleich zur Legalbewährung zuzumessen ist. So kann davon ausgegangen werden, dass Merkmale der Person (wie z.B. soziale Kompetenz, Impulskontrolle, schulische und berufliche Qualifikation) in der Kette der Faktoren, die den Ausstieg aus kriminellem Verhalten begünstigen als direkte Behandlungsziele leichter zu beeinflussen und als Evaluationskriterien zu erfassen sind. Bei alleiniger Untersuchung dieser Merkmale bleibt jedoch offen, ob und in welchem Umfang diese sich auf die tatsächliche Lebenssituation (berufliche Integration, Partnerschaft, Familie, soziales Umfeld) nach Haftentlassung auswirken und wie wiederum diese Lebenssituation kriminelles Verhalten begünstigt oder den Ausstieg aus kriminellen Karrieren erleichtert. Aus Evaluationsperspektive wäre demzufolge die gleichzeitige Betrachtung verschiedener, in unterschiedlichen Stadien der vermuteten Kausalkette von Behandlungswirksamkeit angesiedelter Effektivitätskriterien zu fordern.

In Kapitel 2 wird ein Überblick über den theoretischen und vollzugspraktischen Hintergrund der Behandlungsforschung gegeben, wobei insbesondere Aspekte der Evaluationsforschung, Behandlungswirksamkeit und Rahmenbedingungen von Behandlung im Jugendstrafvollzug berücksichtigt werden. In Kapitel 3 werden drei für die Behandlungsforschung im Strafvollzug relevante Fragestellungen aufgegriffen und anhand eigener empirischer Untersuchungen bearbeitet. Zum einen wird die Problematik der Verwendung psychodiagnostischer Instrumente in der Behandlungsuntersuchung und Forschung mit Inhaftierten untersucht. Zum anderen werden die Realisierbarkeit unterschiedlicher Untersuchungsdesigns in der Strafvollzugsforschung vor dem Hintergrund von Validitätsgefährdungen sowie im dritten Abschnitt die Verwendung unterschiedlicher Effektivitätskriterien von Behandlung erörtert. In Kapitel 4 werden die wesentlichen Befunde zusammengefasst und ihre Bedeutung für den Jugendstrafvollzug und die Behandlungsforschung im Strafvollzug diskutiert.

2 Hintergrund

Den thematischen Hintergrund der Arbeit bilden methodische und praktische Aspekte der Behandlungsforschung und -evaluation im Strafvollzug sowie die konkreten Bedingungen von Behandlung im deutschen Jugendvollzug. Zunächst wird deshalb die Rolle der Evaluationsforschung in der Kriminalprävention und Straftäterbehandlung näher beleuchtet und der Frage nachgegangen, weshalb und in welcher Form kriminalpräventive Programme evaluiert werden sollten. Im Anschluss wird kurz der aktuelle Kenntnisstand der Forschung zur Straftä-

terbehandlung referiert, deren pessimistisches „nothing works“ der 70er Jahre mittlerweile von einem „*what works how with whom under which conditions* [Hervorhebung v. Verf.]“ (Steller, 1994, S. 10) abgelöst wurde. Abschließend werden Ziele und Bedingungen von Behandlung im Jugendstrafvollzug dargestellt, in dem die vorgestellten empirischen Untersuchungen durchgeführt wurden.

2.1 Evaluation in der Kriminalprävention

Nach traditionellem Verständnis richtet sich Präventionsarbeit an Personen, die noch nicht in eine Problemlage geraten sind oder die noch kein Problemverhalten zeigen (MacKenzie, 1997). Ziel dieser primären oder universellen Prävention ist die Reduzierung der Auftretenswahrscheinlichkeit eines unerwünschten, im Falle der Kriminalprävention meist devianten oder kriminellen Verhaltens, beispielsweise durch Aufklärungsmaßnahmen. Gemäß der Zielgruppe und des Interventionszeitpunkts werden weitere Formen unterschieden (vgl. Bliesener, 2008; Lösel, 2004). Sekundäre Prävention umfasst die Früherkennung und Frühbehandlung und richtet sich in der selektiven Form an Risikogruppen, bei denen die Verringerung von Risikofaktoren für deviantes Verhalten oder in der indizierten Form die Besserung bereits bestehenden devianten Verhaltens angestrebt wird. Tertiäre Prävention bezieht sich auf die Vermeidung von Folgeschäden bereits eingetretenen kriminellen Handelns und von Rückfällen, diesem Präventionsbereich sind folglich der Strafvollzug bzw. die Straftäterbehandlung und Resozialisierung zuzuordnen (Bliesener & Köhnken, 2005).

Die Notwendigkeit der Evaluation als wissenschaftlicher Wirksamkeitsüberprüfung kriminalpräventiver Maßnahmen leitet sich nicht nur her aus der Ökonomie einer möglichst effizienten Mittelverwendung, die eine Bevorzugung besonders wirksamer Programme gegenüber weniger wirksamen fordert, sie lässt sich zusätzlich dadurch begründen, dass auch gut intendierte Interventionsmaßnahmen mitunter negative Folgen mit sich bringen können. So erwiesen sich die Familienbegleitung der Cambridge-Somerville Youth Study oder die vermeintlich abschreckenden Scared-Straight-Programme für kriminelle Jugendliche in Evaluationen nicht als unwirksam, sondern als schädlich. So zeigten in der Cambridge-Somerville Youth Study Jugendliche aus Risikofamilien, denen u.a. Unterstützung durch einen Sozialarbeiter geboten und die zu verschiedenen sozialen Aktivitäten motiviert wurden in einer Nachuntersuchung höhere Delinquenzbelastung, geringere Lebenserwartung und mehr psychiatrische Auffälligkeiten im Vergleich zur unbehandelten Kontrollgruppe (McCord, 1992). Für das Scared-Straight-Programm zeigte sich nach der fast flächendeckenden Einführung in den

USA, dass Teilnehmer der Maßnahme eine höhere Delinquenzbelastung aufwiesen als Nichtteilnehmer (McCord, 2003; Petrosino, Turpin-Petrosino & Buehler, 2003).

Seit den 90er Jahren hat sich der maßgeblich von David Farrington und Lawrence Sherman propagierte Ansatz der evidenzbasierten Kriminalprävention etabliert (Graebisch, 2004; Lösel, 2009). Ziel dieses Ansatzes ist die Generierung und Nutzung empirisch fundierter Wirksamkeitsnachweise für kriminalpräventive Maßnahmen nach dem Vorbild medizinischer Forschung. Kennzeichen ist eine quantitative Forschungslogik, deren Untersuchungsdesign idealerweise am experimentellen Kontrollgruppendesign orientiert ist. Maßnahmen werden als erfolgreich klassifiziert, wenn mehrere unabhängige Studien hoher methodischer Güte nachweisen können, dass eine positive Programmwirksamkeit vorliegt (Farrington, Gottfredson, Sherman & Welsh, 2002). Auch wenn die Forderung nach empirischer Begründung politischer Entscheidungen nicht neu ist (vgl. Campbell, 1969; Orr, 1999), so gelangte die evidenzbasierte Kriminalprävention vor allem in den USA und Großbritannien in jüngster Zeit zu bemerkenswerter Popularität, was sich in der Gründung von Einrichtungen wie der Campbell Collaboration im Jahr 2000 zeigt (Marshall, 2005).

Kritisch merkt Heinz (1998, S. 54) zur empirischen Begründung kriminalpräventiver Interventionen an: „Gegenüber Erwartungen, Effekte kriminalpräventiver Maßnahmen seien eindeutig feststellbar, ist angesichts der Komplexität der Zusammenhänge Skepsis angebracht. Kriminalpräventive Projekte sollten sich deshalb nicht in die Falle begeben, ihre Berechtigung von dem empirisch belegbaren Nachweis ihrer Wirkung abhängig zu machen; das Strafrecht macht es auch nicht. Andererseits bedeutet dies nicht den Verzicht auf jegliche Programmevaluierung, insbesondere nicht den Verzicht auf die Kontrolle unbeabsichtigter und unerwünschter Nebenfolgen.“

In Deutschland wird die Mehrheit der durchgeführten kriminalpräventiven Programme bis heute nicht oder nicht hinreichend wissenschaftlich evaluiert. Die Evaluationspraxis ist heterogen (Holthusen & Lüders, 2003), es wird mehr Begleit- als Wirkungsforschung betrieben, fachliche Standards sind nicht etabliert und insbesondere das Kontrollgruppenexperiment findet kaum Verwendung (Coester, Bannenberg & Rössner, 2007). Derzeit unter evidenzbasierten Gesichtspunkten evaluierte Programme betreffen vor allem Verhaltensauffälligkeiten bei Kindern (z.B. Kuschel et al., 2003; Lösel & Beelmann, 2003) und sind der primären oder sekundären Prävention zuzuordnen. Für die Straftäterbehandlung liegen dagegen trotz intensiver Forschung zur Sozialtherapie in den 70er und 80er Jahren bislang keine nach den genannten methodischen Kriterien gesicherte Erkenntnisse über die Wirksamkeit von Interventionsmaßnahmen vor.

2.2 Kenntnisstand der Forschung zur Straftäterbehandlung

Befasst man sich mit Behandlung im Strafvollzug als Teilbereich der Kriminalprävention, so ist trotz Fortschritten der Behandlungsforschung insbesondere auf der Meta-Ebene weiterhin ein Mangel an empirisch abgesichertem Wissen um die Wirksamkeit konkreter Einzelinterventionen zu konstatieren. Deutlich wurde die unklare Befundlage erstmals 1968, als der Soziologe Robert Martinson vom Staat New York beauftragt wurde, eine Übersicht über den Forschungsstand zu rehabilitativen Maßnahmen im Strafvollzug zusammenzustellen. Nach anfänglichen Widerständen gegen die Veröffentlichung erschien eine Zusammenfassung unter dem Titel „What works? Questions and answers about prison reform“ (Martinson, 1974). Die Überblicksarbeit bezog sich auf Maßnahmen wie Betreuung, Psychotherapie, Gruppentherapie sowie schulische, berufliche und medizinische Maßnahmen, die von 1945-1967 in den USA systematisch evaluiert wurden – mit ernüchternden Befunden. In einer Auswertung von 231 Einzelstudien stellte Martinson fest, die Daten gäben insgesamt wenig Grund anzunehmen, dass eine derzeit bekannte Behandlungsmethode die Rückfälligkeit von Straftätern reduzieren würde. Der durch Martinson begründete und bald in weiten Teilen der Kriminalpolitik verbreitete Behandlungspessimismus („nothing works“) ist bei näherer Betrachtung der Ergebnisse jedoch nicht gerechtfertigt, zumal Martinson selbst eine mögliche Erklärung lieferte: „It is just possible that some of our treatment programs are working to some extent, but that our research is so bad that it is incapable of telling“ (Martinson, 1974, S. 49). Weiterhin kann angemerkt werden, dass Martinson eine Auswahl schlecht implementierter und wenig erfolgreicher Projekte als Belege heranzog, seine Literaturübersicht nicht die Möglichkeiten einer Effektivitätsintegration nutzte und sich zudem allein auf das Wirksamkeitskriterium Rückfälligkeit stützte (Sarre, 1999). Auch bot der veröffentlichte Projektbericht (Lipton, Martinson & Wilks, 1975) eine differenziertere Darstellung und Martinson selbst revidierte später seine pessimistische Einschätzung der Behandlungswirksamkeit (Lösel, 1994; Martinson, 1979).

Eine neue Phase der Behandlungsforschung wurde durch die Verwendung metaanalytischer Verfahren eingeleitet (z.B. Andrews et al., 1990; Gendreau, Little & Goggin, 1996; Lipsey & Wilson, 1998; Lösel, Köferl & Weber 1987). Die Methode ermöglicht die Bestimmung der Behandlungswirksamkeit über eine Vielzahl zugrundeliegender Einzelstudien hinweg, die Identifikation besonders wirksamer Behandlungsmaßnahmen bzw. -methoden sowie den Einbezug unterschiedlicher Effektivitätskriterien. Beispielsweise integrierten Lösel et al. (1987) sowie später Lösel (1994) Studien zur deutschen Sozialtherapie und ermittelten eine Effektstärke von $r = .11$ zugunsten der Behandlung. Dies bedeutet, dass Behandelte im Durchschnitt um 11 Prozentpunkte seltener rückfällig werden oder beim Vergleich von Per-

sönlichkeitsmerkmalen besser abschneiden als nichtbehandelte Insassen (Lösel, 1994).¹ Auswertungen für spezifische Straftätergruppen führten zu ähnlichen Ergebnissen – sowohl für Sexualstraftäter (Lösel & Schmucker, 2005) als auch für junge Straftäter (Dowden & Andrews, 1999; Lipsey, 1992). Die durchschnittliche Wirksamkeit der Behandlung bewegt sich dabei nach statistischen Kriterien (Cohen, 1988) im Bereich schwacher Effekte, die aber trotzdem von praktischer Bedeutsamkeit sind.²

Noch wichtiger als der globale Nachweis von Behandlungswirksamkeit ist die Identifikation besonders effektiver Behandlungsmaßnahmen. Zum einen gehören nach den Ergebnissen von Metaanalysen kognitiv-behaviorale Therapieformen zu den besonders wirksamen Behandlungsmethoden (z.B. Landenberger & Lipsey, 2005), zum anderen lassen sich aus ihnen allgemeine Prinzipien der Straftäterbehandlung ableiten. So erwiesen sich Behandlungsmaßnahmen als besonders erfolgreich, wenn sie dem Risiko-, Bedürfnis- und Ansprechbarkeitsprinzip Rechnung tragen (Andrews et al., 1990). Das Risikoprinzip besagt, dass das Ausmaß der Behandlung an das individuelle Rückfallrisiko gekoppelt sein sollte, das Bedürfnisprinzip fordert die Ausrichtung der Behandlung an individuellen Risikofaktoren bzw. Defiziten und das Ansprechbarkeitsprinzip die Berücksichtigung der Fähigkeiten und Lerngewohnheiten der Straftäter. Insgesamt verschiebt sich die Forschungsfrage der letzten Jahrzehnte damit also von der Wirksamkeit von Straftäterbehandlung als solcher hin zu der Frage, wie genau Behandlung im Strafvollzug wirkt und welche Maßnahmen bei welchen Straftätern unter welchen Bedingungen effektiv sind (Cullen & Gendreau, 2001; Steller, 1994).

Trotz des Nachweises von Behandlungseffektivität auf der Meta-Ebene gibt es bezogen auf konkrete Maßnahmen weiteren Forschungsbedarf. Ein Beispiel ist der ungeklärte Wirksamkeitsstatus des Anti-Aggressivitäts-Trainings, einem kognitiv-behavioralen Gruppentraining, das in seiner ursprünglichen Konzeption vor allem durch das konfrontative Element des „heißen Stuhls“ bekannt geworden ist. Obwohl die Befunde zur Effektivität uneinheitlich sind, hat die Maßnahme mittlerweile weite Verbreitung im Jugendstrafvollzug (Bosold, Prasse & Lauterbach, 2006) und der ambulanten Jugendhilfe gefunden. Einigen positiven Befunden zur Wirksamkeit – meist gewonnen in Selbstevaluationen der Programmentwickler bzw. Programmdurchführenden im Jugendstrafvollzug (Weidner, 1990; Wolters, 1992) sowie spä-

¹ Bei der Verwendung psychologischer Maße als Erfolgskriterien finden sich im Allgemeinen stärkere Effekte (z.B. Lipsey & Wilson, 1998; Grietens & Hellinckx, 2004). Dies ist insofern nachvollziehbar, da es sich im Vergleich zum Rückfall um das proximalere Maß handelt.

² Da in Metaanalysen zur Straftäterbehandlung häufig hohe Varianzanteile auf die Qualität des Studiendesigns zurückzuführen sind, raten einige Autoren an, resultierende Effektstärken entsprechend vorsichtig zu interpretieren (z.B. Endrass, Rossegger, Noll & Urbaniok, 2008). Selbst bei Berücksichtigung der Designqualität besteht ein weiteres Problem darin, dass Moderatoreffekte korreliert sein können, so dass schwer entscheidbar ist, ob beispielsweise die Qualität des Studiendesigns oder der Implementation einer Maßnahme das Ergebnis beeinflusst (Lipsey, 2003).

ter im ambulanten Bereich (Rau, 2006a, 2006b; Schanzenbächer, 2003) – stehen vorsichtigere Einschätzungen unabhängiger Evaluationsforscher gegenüber (Ohlemacher, Sögding, Höynck, Ethé & Welte, 2001; Weichold, 2004). Abgesehen von möglichen Interessenkonflikten bei Selbstevaluationen³ kann die Heterogenität der Befunde auch auf die Güte der verwendeten Evaluationsdesigns und unterschiedlich gewählte Wirksamkeitskriterien zurückgeführt werden.

2.3 Behandlung im Jugendstrafvollzug

Die Jugendstrafe wird in Deutschland in 28 Jugendstrafanstalten vollzogen. Im März 2008 verbüßten 6293 männliche und 264 weibliche Gefangene eine Haft, davon mehr als 90 % im geschlossenen Vollzug (Statistisches Bundesamt, 2009). Der Jugendstrafvollzug unterscheidet sich vom Erwachsenenstrafvollzug durch seinen Erziehungsauftrag (Walter, 2002), der in § 2 JGG festgeschrieben ist und in den Jugendstrafvollzugsgesetzen der Länder aufgegriffen wurde.⁴ Kennzeichen des Jugendstrafvollzugs ist folglich seine Erziehungs- und Behandlungsorientierung, wobei von der Annahme ausgegangen wird, dass junge Menschen durch Behandlung leichter zu beeinflussen sind als Erwachsene. Dabei ist Behandlung im Jugendstrafvollzug nicht zu reduzieren auf sozial- und psychotherapeutische Maßnahmen, sie erstreckt sich auf schulische und berufliche Förderung, Förderung sozialer Kompetenzen, Freizeitangebote sowie deliktspezifische Angebote und ergänzende Hilfen (Basqué, 1994; Hosser & Bosold, 2008). Im Bereich der psychotherapeutischen Behandlung bezieht sich der Behandlungsbegriff dabei in erster Linie auf direkt oder indirekt rückfallpräventive Interventionen und nicht auf Behandlung im Rahmen einer psychotherapeutischen Grundversorgung (vgl. Behnke & Endres, 2008).

Vor dem Hintergrund, dass beispielsweise in Nordrhein-Westfalen zwei Drittel der Jugendstrafgefangenen vor Inhaftierung ohne Schulabschluss und fast 90 % ohne Berufsausbildung sind (Wirth, 2006b) kommt der schulischen und beruflichen Bildung im Jugendstrafvollzug große Bedeutung zu. Entsprechend gehören Haupt- und Realschulkurse, Berufsausbildungen in meist handwerklichen aber auch technischen Berufen sowie berufsvorbereitende

³ Zu möglichen Interessenkonflikten bei Selbstevaluationen siehe Müller-Kohlenberg (2006). Empirische Hinweise finden sich in Metaanalysen, die höhere Effektstärken für Untersuchungen berichten, bei denen Evaluierende auch Maßnahmendurchführende sind (Lipsey, 1995; Petrosino & Soydan, 2005). Neben einer selektiven Ergebnisdarstellung und Versuchsleitereffekten wird als Erklärung aber auch eine bessere Maßnahmenintegrität diskutiert, wenn Forscher an der Entwicklung oder Implementation beteiligt sind (Lipsey, 1995).

⁴ Beispielsweise gilt für Niedersachsen „Der Vollzug ist erzieherisch zu gestalten. Zur Erreichung des Vollzugszieles [...] ist die oder der Gefangene in der Entwicklung von Fähigkeiten und Fertigkeiten sowie der Bereitschaft zu einer eigenverantwortlichen und gemeinschaftsfähigen Lebensführung in Achtung der Rechte anderer zu fördern.“ (§ 114 NJVollzG). Ähnliche Formulierungen finden sich in anderen Ländergesetzen (vgl. Höynck et al., 2008).

Maßnahmen zum Standardangebot der Jugendanstalten. Wurde noch vor einiger Zeit bemängelt, dass das berufserziehende Angebot der Anstalten quantitativ und qualitativ unzureichend sei (Stentzel, 1990), so wird mittlerweile dem Bildungsangebot im Jugendstrafvollzug ein hoher Standard bescheinigt (Wirth, 2002). Jedoch erleben Inhaftierte die Zuweisung zu Maßnahmen häufig als wenig an den eigenen Interessen ausgerichtet, was sich negativ auf die Teilnahmemotivation auswirken kann (Bereswill, Koesling & Neuber, 2007). Zudem ist die Beendigung umfangreicherer Ausbildungsmaßnahmen nicht immer mit der Haftdauer vereinbar, weshalb in Modellprojekten modularisierte Ausbildungen angeboten werden (Klein, Petran & Weber, 2007).

Deliktspezifische Behandlungsmaßnahmen werden in erster Linie für Gewalt- und Sexualstraftäter angeboten, störungsspezifische Angebote richten sich an Inhaftierte mit Alkohol- oder Drogenproblemen (Repp, Köhler & Hinrichs, 2004). Verbreitete Behandlungsprogramme sind das Anti-Aggressivitäts-Training zur Reduktion der Gewaltneigung (Heilemann & Fischwasser-von Proeck, 2001) sowie das soziale Training zur Förderung sozialer Kompetenzen und Vermittlung lebenspraktischer Hilfen (Otto, 1994). Sexualstraftäter werden oft – analog zum Erwachsenenvollzug – in eigenen Abteilungen behandelt. Neben Gruppenangeboten existieren Einzelfallhilfen wie Krisenintervention, Einzelpsychotherapie und unterschiedliche Beratungsangebote (z.B. Schuldnerberatung, Berufsberatung und Seelsorge). Besondere Problemgruppen im Jugendstrafvollzug stellen Ausländer und Spätaussiedler dar (Walter, 2003; Wirth, 1998), die oft aufgrund von Sprachbarrieren von Förder- und Behandlungsmaßnahmen ausgeschlossen sind oder sich diesen aufgrund ihrer subkulturellen Einbindung verschließen. Allerdings werden auch Förderprogramme für gerade diese Zielgruppen angeboten, wie z.B. Deutsch-Förderkurse und Integrationsmaßnahmen.

Welche Form von Behandlung und Förderung ein Inhaftierter erhält, wird im Erziehungs- und Behandlungsplan festgehalten, der auf Grundlage der Behandlungsuntersuchung in den ersten Wochen der Haftstrafe erstellt und im Verlauf der Haft kontinuierlich fortgeschrieben wird (Müller, 2002). Auf Basis von Aktdaten, persönlicher Exploration individueller Entwicklungsaspekte, Verhaltensbeobachtungen und standardisierter Verfahren (Intelligenz- und Persönlichkeitstests sowie Rückfallprognoseinstrumente) werden ein Erklärungsmodell der Delinquenz und eine Legalprognose erstellt (Cottonaro & Suhling, 2007). Inhaftierte mit ungünstiger Prognose werden als besonders behandlungsbedürftig angesehen und entsprechend ihrer Defizite einzelnen Behandlungsmaßnahmen zugewiesen. Als weitere Kriterien der Zuweisung werden neben der Bedürftigkeit auch die Behandlungsfähigkeit und Behandlungsmotivation berücksichtigt (Behnke, 2004; Bosold, 2008). Zur Behandlungsfähigkeit von Inhaf-

tierten wird kritisch angemerkt, dass nicht eine selektive Strategie der Auswahl geeigneter Teilnehmer für ein bestehendes Therapieangebot sondern eine adaptive Strategie der Auswahl der geeigneten Maßnahme bei gegebenem Störungsbild des Inhaftierten verfolgt werden sollte (Dahle, 1997), was allerdings in der strafvollzuglichen Praxis noch zu wenig berücksichtigt wird. Eine Besonderheit der Behandlungsmotivation von Straftätern ist, dass diese häufig keinen Leidensdruck verspüren und demzufolge nicht intrinsisch zu einer Behandlungsteilnahme motiviert sind. Nach Dahle ist intrinsische Motivation jedoch keine Voraussetzung von Therapie im Strafvollzug, sondern eines ihrer Therapieziele. Von Bedeutung ist in diesem Zusammenhang, dass Motivation kein statisches Merkmal ist, sondern sich eine extrinsisch motivierte Behandlungsteilnahme (z.B. durch in Aussicht gestellte Lockerungen) im Verlauf der Behandlung zu einer intrinsisch am Behandlungsziel orientierten Motivation verändern kann (Dahle, 1997; Suhling & Cottonaro, 2005).

Mitunter entspricht die Popularität einer Behandlungsmaßnahme nicht ihrer Bedeutung im Sinne der tatsächlich erreichten Behandlungsquote. Beispielsweise wird in fast allen Jugendanstalten das bereits erwähnte Anti-Aggressivitäts-Training angeboten, jedoch werden meist weniger als 20 Insassen pro Jahr und Anstalt im Rahmen dieser Maßnahme behandelt. Vor dem Hintergrund, dass etwa 48 % der Inhaftierten ein Gewaltdelikt (ohne Sexualdelikte) begangen haben (Dünkel & Geng, 2007) ist die mit Anti-Aggressivitäts-Trainings realisierte Behandlungsquote vergleichsweise gering, selbst wenn man berücksichtigt, dass nicht alle Gewaltstraftäter eine Indikation für das Training aufweisen. Für psychotherapeutische Behandlung im Strafvollzug insgesamt wird angenommen, dass ein erheblicher Anteil von etwa 35-50 % der Inhaftierten behandlungsbedürftig ist und der Bedarf deutlich über dem derzeitigen Therapieangebot liegt (Dahle, 1997). Umso wichtiger ist es, Behandlung nicht nach dem Gießkannenprinzip durchzuführen, sondern die Inhaftierten gezielt solchen Maßnahmen zuzuführen, von denen sie in besonderem Maße profitieren – was nur durch eine reliable und valide Straftäterdiagnostik gewährleistet werden kann.

3 Eigene empirische Untersuchungen

Aus dem Vorangegangenen ergibt sich ein Forschungsdefizit in den Bereichen der Straftäterdiagnostik und der Wirksamkeit strafvollzuglicher Behandlung. Die zu Inhaftierungsbeginn durchgeführte Behandlungsuntersuchung im Strafvollzug stellt eine Weichenstellung dar, die von erheblicher Bedeutung für den Vollzugsverlauf und den Erfolg von Behandlung im Strafvollzug ist. Da in dieser Untersuchung unter anderem auf standardisierte Instrumente zurückgegriffen wird, deren Validität für die untersuchte Population nicht gesichert ist, besteht ein

möglicher Schritt der verbesserten Behandlungszuweisung in der Überprüfung und gegebenenfalls der Revision bestehender Instrumente. Weiterhin steht für viele der im Jugendstrafvollzug angebotenen Behandlungsprogramme ein methodisch stringent geführter Wirksamkeitsnachweis noch aus, unter anderem weil vollzugspraktische und rechtliche Gegebenheiten besondere Schwierigkeiten für die Evaluationsforschung mit sich bringen. Vor diesem Hintergrund werden drei methodische Fragestellungen näher untersucht, die für den Strafvollzug und die Strafvollzugsforschung von hoher praktischer Relevanz sind.

1. Sind für die Allgemeinpopulation entwickelte standardisierte psychodiagnostische Instrumente für den Einsatz mit jungen Straftätern z.B. in der Behandlungsuntersuchung geeignet und welche Einschränkungen müssen in Abhängigkeit vom Anwendungskontext gegebenenfalls in Kauf genommen werden?
2. Sind Forderungen an die Güte von Untersuchungsdesigns z.B. seitens Vertretern einer evidenzbasierten Kriminalprävention in der strafvollzuglichen Behandlungsforschung realisierbar und welche Konsequenzen ergeben sich aus der Verwendung von Designs geringerer Güte?
3. Welche Kriterien der Behandlungseffektivität sollten über das üblicherweise verwendete Legalbewährungskriterium hinaus verwendet werden und wie stehen diese in Bezug zum Legalverhalten?

Empirische Grundlage zur Beantwortung dieser Fragen sind Ergebnisse aus verschiedenen Einzelprojekten des Kriminologischen Forschungsinstituts Niedersachsen (KFN). Das darunter umfangreichste ist die Längsschnittuntersuchung „Entwicklungsfolgen der Jugendstrafe“ (Greve, 2004; Hosser & Bosold, 2006), die von der Deutschen Forschungsgemeinschaft gefördert wurde und an das Vorgängerprojekt „Gefängnis und die Folgen“ (Greve & Hosser, 1998; Greve, Hosser & Pfeiffer, 1997) anschließt. Untersucht wird die Bedeutung einer Jugendstrafe als „Entwicklungsintervention“ und dabei insbesondere, welche Haftbedingungen und Behandlungsmaßnahmen den Erwerb formaler Qualifikation und sozialer Kompetenzen fördern und wie die Haft und die durchgeführten Maßnahmen sich auf das spätere Legal- und Sozialverhalten auswirken. Im Rahmen der Untersuchung wurden 2405 männliche deutsche Erstinhaftierte aus sechs norddeutschen Jugend- und Jungtäteranstalten mehrfach während der Haft und nach Haftentlassung standardisiert befragt und ergänzend eine Analyse der Gefangenepersonalakten und der Einträge im Bundeszentralregister durchgeführt.

Drei weitere Projekte entstanden aus Kooperationen des KFN mit der Jugendanstalt Hameln. Zwei Untersuchungen befassen sich mit der Evaluation delikt-spezifischer Behandlungsmaßnahmen der Jugendanstalt. Zum einen wurde das Training „Leben ohne Gewalt or-

ganisieren“ (LoGo), einer Weiterentwicklung des ursprünglich in Hameln entwickelten Anti-Aggressivitäts-Trainings, auf seine Wirksamkeit hin untersucht (Bosold & Lauterbach, 2010; Bosold, Lauterbach & Hosser, 2006). Im Vergleich zu einer Kontrollgruppe wurden bei 49 Trainingsteilnehmern Behandlungseffekte in den Bereichen Aggressivitätsneigung, Neutralisierungstendenzen und Ressourcenstärkung sowie Rückfalldaten überprüft. Zum anderen wurde die Behandlung junger Sexualstraftäter in der Sozialtherapeutischen Abteilung II evaluiert (Hosser, Bosold & Lauterbach, 2006). Im Rahmen der Untersuchung wurden bei 17 Insassen im Prä-Post-Vergleich Veränderungen deliktspezifischer Einstellungsmaße und personaler Ressourcen erhoben. Eine weitere Studie (Hosser, Lauterbach & Camehn, 2008) befasst sich mit der Auswertung von 775 im Rahmen der Eingangsdagnostik der Jugendanstalt mittels des Freiburger Persönlichkeitsinventars (FPI-R; Fahrenberg, Hampel & Selg, 2001) durchgeführten Interviews. Die erhobenen Persönlichkeitsdaten wurden auf ihre Validität für die spezifische Population Jugendstrafgefangener, insbesondere interne Konsistenz, faktorielle Validität und die Einhaltung altersgerechter Normwerte überprüft.

3.1 Einsatz standardisierter Persönlichkeitstests in der Straftäterdiagnostik

Die Behandlungsuntersuchung zu Beginn einer Haftstrafe dient der Untersuchung der Persönlichkeit und der Lebensverhältnisse der Gefangenen als Voraussetzung der Vollzugs- und Behandlungsplanung (Laubenthal, 2003; Stock, 1993). Die Erfassung von Persönlichkeitsmerkmalen geschieht in der Regel durch standardisierte Breitband-Persönlichkeitstests wie beispielsweise das Freiburger Persönlichkeitsinventar (FPI-R; Fahrenberg, Hampel & Selg, 2001) oder auf einzelne Persönlichkeitsaspekte bezogene Instrumente wie den Fragebogen zu Kompetenz- und Kontrollüberzeugungen (FKK; Krampen, 1991).⁵

Grundannahme der Berücksichtigung der Täterpersönlichkeit ist, dass neben der sozialen Situation spezifische Ausprägungen und Konstellationen von Persönlichkeitsmerkmalen Mitursache des abweichenden Verhaltens des Straftäters darstellen und ein Einwirken auf diese individuellen Dispositionen durch Behandlung zur Vorbeugung einer erneuten Straffälligkeit beiträgt (Lösel, 1975, 1983). Beispielsweise werden Persönlichkeitsmerkmale wie Selbstkontrolle (Mischel, 1961), Extraversion (Eysenck, 1965), Kontrollüberzeugungen (Scheurer, 1993) und Empathie (Miller & Eisenberg, 1988) als Verhaltensdispositionen aufgefasst, die delinquentes Verhalten begünstigen oder diesem entgegenwirken (vgl. Ullrich, 1999). Bereits

⁵ In dieser Arbeit mit dem Terminus psychodiagnostische Instrumente ausdrücklich nicht gemeint sind Prognoseinstrumente wie z.B. die Psychopathy Checklist (PCL-R), das Level of Service Inventory (LSI-R) oder das Historical-Clinical-Risk Management Schema (HCR-20), die auf der Basis von Biografie, krimineller Vorgeschichte sowie verschiedener personaler Risiko- und Schutzfaktoren eine Bewertung der Rückfallwahrscheinlichkeit erlauben (vgl. Dahle, 2007; Dahle, Schneider & Ziethen, 2007; Nedopil, 2006).

im psychopathologischen Bereich liegt zudem das Psychopathiekonzept (Hare, 1980), dessen zentrale Merkmale neben der Verhaltensauffälligkeit in Gefühlskälte und Gewissenlosigkeit bestehen.

Die erhobenen Persönlichkeitsdaten werden bei Entscheidungen bezüglich des Haftverlaufs, der Gewährung von Lockerungen und der Zuweisung zu einzelnen Behandlungsangeboten herangezogen. Da die Entscheidungen zwar nicht ausschließlich auf Grundlage, jedoch unter Zuhilfenahme standardisierter Testverfahren getroffen werden, ist die Validität der verwendeten Instrumente von erheblicher praktischer Bedeutung für die Insassen und die Effektivität strafvollzuglicher Behandlung. Weiterhin kommen standardisierte psychodiagnostische Instrumente nicht nur in der Eingangsuntersuchung, sondern auch im Rahmen von Forschungs- und Evaluationsvorhaben zum Einsatz (z.B. Kury & Beckers, 1983; Ortmann, 1987; Weidner & Wolters, 1991), wo kriminalitätsrelevante Einstellungs- und Persönlichkeitsmerkmale als Kriterien des Erfolgs oder Misserfolgs von Maßnahmen herangezogen werden.

Probleme der Anwendung standardisierter Persönlichkeitsinventare im Strafvollzug liegen darin begründet, dass die wenigsten psychodiagnostischen Instrumente speziell für Straftäter entwickelt wurden. Ausnahmen sind der Prisonisierungsfragebogen von Ortmann (1987) oder der Persönlichkeitsfragebogen für Inhaftierte (PFI; Seitz, 1983), diese beziehen sich allerdings weniger auf Aspekte der Persönlichkeit als auf Haftsituation und Hafterleben und sind daher für die Behandlungsuntersuchung zu Haftbeginn kaum geeignet. Die meisten anderen eingesetzten Instrumente wurden für die Allgemeinpopulation entwickelt und eine Übertragbarkeit auf die spezifische Personengruppe von Straftätern ist demzufolge nicht ohne weiteres zulässig (z.B. Heilbrun, 1992; Littmann, 1992). Fraglich ist bei mehrdimensional faktorenanalytisch entwickelten Instrumenten die faktorielle Validität beim Einsatz mit Straftätern und in der Folge die Gültigkeit und Abgrenzbarkeit der erhobenen Konstrukte (Kury, 1983b). Weiterhin können Verständnisprobleme aufgrund hoher sprachlicher und kognitiver Anforderungen (Stieglitz, 1993) sowie fehlendem Bezug zur Lebenswelt der Befragten („Mittelschichtorientierung“; Kury & Beckers, 1983) die Validität der Ergebnisse negativ beeinflussen. Als weitere Kritikpunkte werden fehlende oder nicht hinreichend aktuelle Normwerte sowie die hohe Zahl von Fehlklassifikationen bei der Risikovorhersage auf Basis von Persönlichkeitsdaten angemerkt (Gudjonsson, 2001).

Neben der Eignung der Instrumente für Straftäter ist die Situation zu berücksichtigen, in der die Testung vorgenommen wird. Anders als im Beratungskontext, wo bei den Klienten ein Eigeninteresse an einer genauen und validen Messung angenommen werden kann, findet die Erhebung bei Inhaftierten in einem Zwangskontext statt. Die Befragten können der Testung

entweder desinteressiert gegenüber stehen oder aber zu Recht annehmen, dass aus den Testergebnissen konkrete den weiteren Haftverlauf betreffende Entscheidungen abgeleitet werden. Mögliche Antworttendenzen sind demzufolge willkürliches Antwortverhalten, Tendenz zur sozialen Erwünschtheit oder gezielte Verfälschung (Kury, 2002). Da Persönlichkeitsinventare für die Befragten hochgradig durchschaubar sind (Krahé, 1989; Kubinger, 2002), sind diese in der Lage, Vermutungen über den Zweck der Erhebung anzustellen und sich im Extremfall bei Untersuchungen zur Schuldfähigkeit als besonders belastet und bei Untersuchungen anlässlich bevorstehender Lockerungen als besonders kompetent darzustellen (Scheurer & Richter, 2005).

Um Antworttendenzen zu begegnen sind in vielen Instrumenten Offenheits-, Erwünschtheits- oder Lügenskalen integriert, mit deren Hilfe unehrliches Antwortverhalten aufgedeckt werden soll. Die Skalen beruhen auf dem Prinzip, relativ verbreitete geringfügige Normverletzungen oder aber wenig verbreitete sozial erwünschte Verhaltensweisen zu erfragen und von der ehrlichen Beantwortung dieser Items auf das Antwortverhalten in anderen Merkmalsbereichen zu schließen (Biedermann, 2007; Seiwald, 2003). Verneint eine Person, zu Hause weniger gute Tischmanieren zu haben als im Restaurant oder hin und wieder ein wenig schadenfroh zu sein (so Items des FPI), so wird sie nach dieser Logik auch Fragen zu anderen unerwünschten Eigenschaften nicht wahrheitsgemäß beantworten und ist von der Auswertung auszuschließen. Zu beachten ist hierbei, dass zwar mit Offenheitsskalen tatsächlich Personen identifiziert werden können, die in übertriebenem Maße sozial erwünscht antworten, dass jedoch der Umkehrschluss unzulässig ist, alle anderen Befragten würden kein sozial erwünschtes Antwortverhalten aufweisen.⁶ In der Konsequenz bieten Offenheits- oder Erwünschtheitsskalen keine Gewähr, dass mit Persönlichkeitsinventaren erfasste Eigenschaften frei von sozialer Erwünschtheit und Verfälschung sind.

Zusammenfassend betrachtet stellt sich die Frage, ob standardisierte für die Normalpopulation entwickelte psychodiagnostische Instrumente für den Einsatz mit Straftätern geeignet sind und im jeweiligen Anwendungskontext valide Ergebnisse erbringen. Dies soll im Folgenden anhand des in der Behandlungsuntersuchung verwendeten Persönlichkeitsfragebogens sowie einer verbreiteten Skala zur Erfassung von Empathie überprüft werden.

⁶ Zwei Argumente können nach Biedermann (2007) hier angeführt werden. Einerseits die Kontextspezifität: Das Bekennen von Normverletzungen im Alltagsbereich ist nicht zwingend ein Prädiktor für ehrliche Beantwortung von Items in spezifischen Bereichen. Insbesondere Straftäter könnten Fragen nach alltäglichen Normverletzungen ehrlich beantworten, aber im Bereich der Delinquenz zu Antwortverfälschungen neigen. Zum anderen variieren Normverletzungen im Grad der Schwere: Alltägliche Normverletzungen sind vergleichsweise unbedeutend, die in anderen Bereichen, insbesondere der Delinquenz dagegen schwerwiegend, weil moralisch negativ belegt und strafrechtlich verfolgt. Befragte könnten also dazu neigen, geringfügige Normverletzungen zuzugeben und bei schwerwiegenderen Normverletzungen zu verfälschen.

3.1.1 Einsatz des FPI-R bei Inhaftierten

Das im Rahmen der Eingangsdiagnostik verwendete FPI-R ist ein mehrdimensionaler Persönlichkeitsfragebogen, der die zehn Dimensionen Lebenszufriedenheit, Soziale Orientierung, Leistungsorientierung, Gehemmtheit, Erregbarkeit, Aggressivität, Beanspruchung, Körperliche Beschwerden, Gesundheitsorgen und Offenheit sowie die zwei Zusatzskalen Extraversion und Emotionalität umfasst. Bei der Überprüfung der psychometrischen Eigenschaften und der faktoriellen Struktur des Instruments konnte mittels konfirmatorischer Faktorenanalysen nur ein ungenügender Modellfit für das zehnfaktorielle Strukturmodell nachgewiesen werden (Hosser, Lauterbach & Camehn, 2008). Dies deutet darauf hin, dass die einzelnen Skalendimensionen schlecht voneinander abgrenzbar und folglich die erhobenen Konstrukte von eingeschränkter Validität sind. Methodenkritisch muss hier angemerkt werden, dass konfirmatorische Faktorenanalysen komplexer Persönlichkeitsinventare häufig zu unbefriedigenden Ergebnissen führen (Borkenau & Ostendorf, 1990; Church & Burke, 1994; Raykov, 1998) da die überprüfte Eindimensionalität der Teilskalen möglicherweise ein zu restriktives Kriterium darstellt. Ungünstige Reliabilitätskennwerte verweisen jedoch ebenfalls auf eine eingeschränkte Skalenqualität.

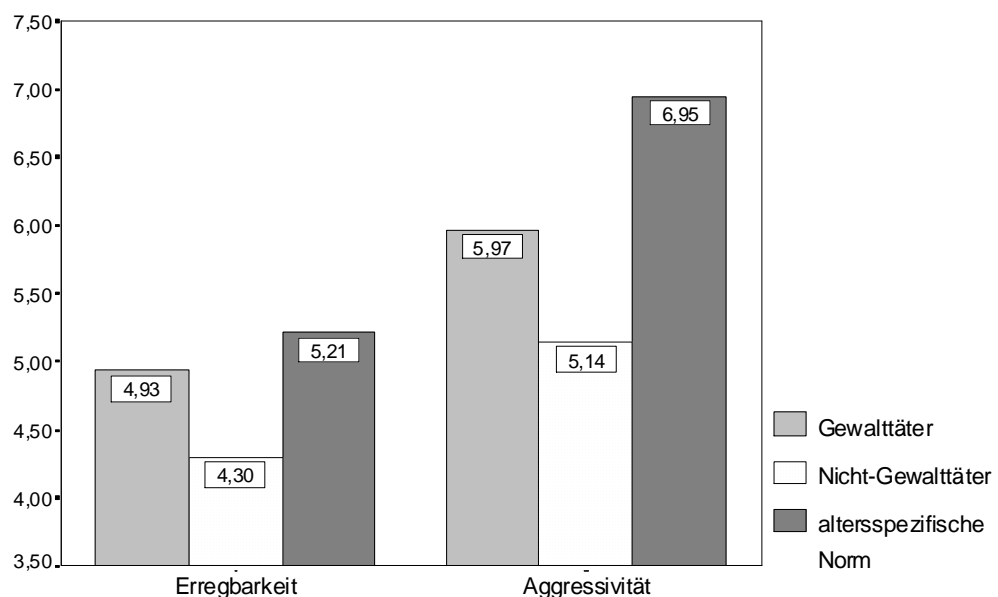


Abbildung 1: FPI-R Skalenwerte im Vergleich zur Norm
(Normwerte aus Fahrenberg, Hampel & Selg, 2001)

Weiterhin führten Gruppenvergleiche bezogen auf die Konstruktvalidität von Teilskalen zu widersprüchlichen Befunden. Einerseits differenzieren die Subskalen Erregbarkeit und Aggressivität hypothesenkonform Gewaltstraftäter von Nicht-Gewaltstraftägern (siehe Abbildung 1). Andererseits erweisen sich die Befragten im Vergleich zur altersspezifischen Norm als we-

niger aggressiv, was kaum anders als durch die Erhebungssituation bedingte Verfälschungsin-tentionen im Sinne eines „faking good“ zu erklären ist. Die Konsequenz dieser Befunde muss lauten, dass das Instrument in seiner jetzigen Form als Entscheidungshilfe für die Zuweisung zu erfolgsversprechenden Behandlungsmaßnahmen nur sehr eingeschränkt verwendet werden sollte. Eine Anpassung des Instruments an die Zielgruppe durch veränderte Itemformulierung mag zwar die faktorielle Validität verbessern, löst jedoch nicht das grundsätzliche Problem der Erhebung von Persönlichkeitsdaten in einem Zwangskontext, insbesondere wenn aus ihnen praktische Konsequenzen abgeleitet werden.

3.1.2 Erfassung von Empathie bei Inhaftierten

Empathie kann unterschieden werden in affektives Mitfühlen oder Nachempfinden des psychischen Zustands einer fremden Person oder kognitives „sich Hineinversetzen“ in deren Lage (Eisenberg & Strayer, 1987). Dabei gelten empathische Fähigkeiten als Ressource, die der Ausübung delinquenten Verhaltens, insbesondere der Gewaltkriminalität entgegenwirkt (Jolliffe & Farrington, 2004; Kerner, 2004). Im Rahmen des Projekts „Entwicklungsfolgen der Jugendstrafe“ wurde Empathie bei 839 Teilnehmern mittels des Interpersonal Reactivity Index (IRI; Davis, 1980, 1983) in der deutschen Version von Enzmann (1996) erhoben. Neben empathischer Anteilnahme und Perspektivübernahme als affektive und kognitive Empathiefacetten werden mit dem Instrument empathischer Distress als negative Reaktion auf die wahrgenommene Notlage anderer, sowie empathische Fantasie erhoben, die sich auf fiktive anstatt realer Personen und Situationen bezieht. Grundsätzlich erfasst der IRI dabei Tendenzen bzw. Neigungen und nicht empathische Fähigkeiten oder Kapazitäten (Davis & Franzoi, 1991).

Die Befunde zeigen, dass das Instrument in seiner ursprünglichen Version nur bedingt für junge Straftäter verwendet werden kann (Lauterbach & Hosser, 2007). Insbesondere die Reliabilität und faktorielle Validität des Instruments erwiesen sich als eingeschränkt, so lange nicht negativ formulierte Items aus den Skalen entfernt wurden. Als Ursache hierfür konnten sprachliche und kognitive Fähigkeiten der Untersuchungsteilnehmer, gemessen an der Schulnote im Deutschunterricht und dem IQ identifiziert werden. Nach einer Revision des Instruments und der Entfernung negativer Items verbesserten sich die faktorielle Validität und Reliabilität. Darüber hinaus konnten Gewaltstraftäter von Nicht-Gewaltstraftätern differenziert und ein nach der Haftentlassung stattgefundener Gewaltrückfall mittels der Subskala Perspektivübernahme signifikant vorhergesagt werden.

Anders als beim FPI-R liegen für das Instrument in seiner deutschen Version keine Vergleichswerte vor, an denen überprüft werden könnte, ob die Angaben der Inhaftierten von der

Normalpopulation abweichen. Hinweise auf eine gezielte Verfälschung lassen sich deshalb aus den Ergebnissen nicht ableiten, allerdings ist diese bei der Erhebung im Rahmen unabhängiger Forschung im Strafvollzug auch weniger zu befürchten. Aufgrund der genannten Einschränkungen muss allerdings eine sprachliche Überarbeitung des Instruments für den Einsatz mit Inhaftierten dringend empfohlen werden.

3.2 Güte von Untersuchungsdesigns in der Behandlungsforschung

In den neugestalteten (Jugend-)Strafvollzugsgesetzen der Länder wurde ein Evaluations- und Forschungsauftrag festgeschrieben. Für Niedersachsen gilt beispielsweise, dass „Erfolge und Misserfolge des Vollzuges, insbesondere im Hinblick auf Rückfallhäufigkeiten“ festzustellen und die „hierfür verantwortlichen Faktoren“ zu erforschen sind (§ 189 Abs. 2, NJVollzG). Eine Schwierigkeit der Behandlungsforschung im Strafvollzug besteht jedoch darin, dass sich übliche Standards der Evaluationsforschung (wie z.B. die Rekrutierung adäquater Vergleichsgruppen; Suhling, 2006) nur sehr eingeschränkt realisieren lassen, was bereits seit der Debatte um die Evaluation der in den 70er Jahren errichteten sozialtherapeutischen Anstalten bekannt ist (vgl. Kury, 1983a).

Einen Maßstab der methodischen Qualität von Evaluationsstudien stellt die Maryland Scientific Methods Scale (Sherman et al., 1997) dar. Fünf Stufen beschreiben das Niveau interner Validität des Untersuchungsdesigns unter Berücksichtigung insbesondere des Kriteriums der Kontrolle von Störvariablen (siehe Abbildung 2). Die Skala reicht von rein korrelativen Zusammenhängen zwischen Interventionsmaßnahme und der abhängigen Variable (Stufe 1) über Prä-Post-Untersuchungen oder Kontrollgruppendesigns ohne nachgewiesene Vergleichbarkeit der Gruppen (Stufe 2), Kontrollgruppenuntersuchungen mit vergleichbarer Kontrollgruppe (Stufe 3), Kontrollgruppenuntersuchungen mit vergleichbarer Kontrollgruppe unter Berücksichtigung weiterer Einflussfaktoren (Stufe 4) bis hin zu randomisierten Kontrollgruppendesigns (Stufe 5).

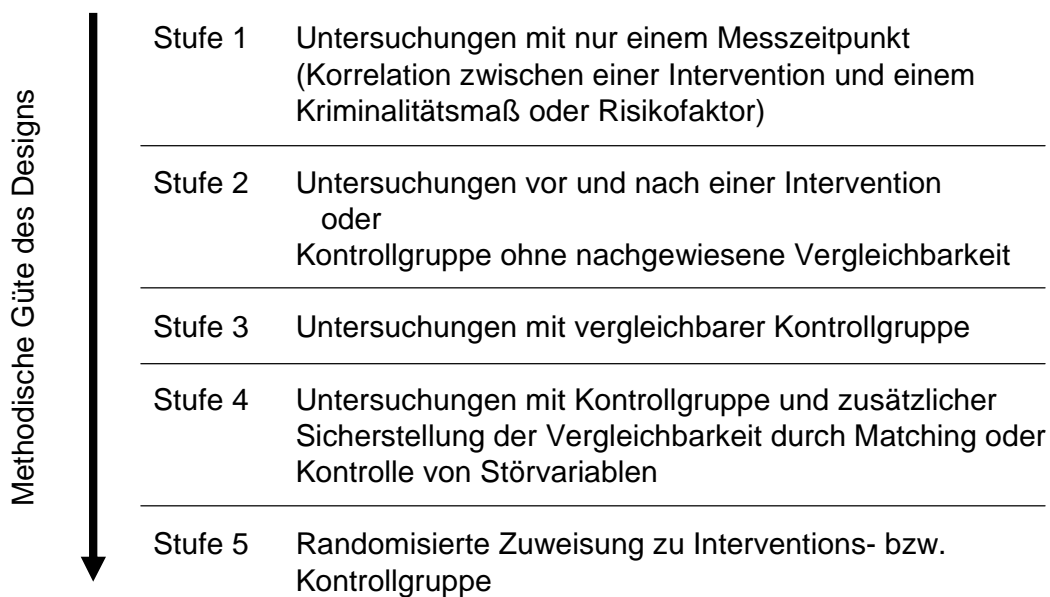


Abbildung 2: Maryland Scientific Methods Scale nach Sherman et al. (1997)

Das in der Maryland-Scale favorisierte randomisierte Kontrollgruppendesign stellt als einziges Untersuchungsdesign sicher, dass Behandlungs- und Kontrollgruppe nicht nur bezogen auf bekannte, sondern auch auf unbekannte konfundierende Variablen äquivalent sind (Boruch, 1997; Lösel & Nowack, 1987). Somit sind im Posttest vorgefundene Unterschiede allein auf die Behandlung zurückzuführen, wenn nicht Probleme der Implementation die interne Validität gefährden. Allerdings erreichen Evaluationsstudien im deutschen Strafvollzug selten das höchste Niveau dieser Skala, da ethische, rechtliche und vollzugspraktische Bedenken einer randomisierten Behandlungszuweisung entgegenstehen. Ausnahmen stellen die Studien von Rasch und Kühl (1977, 1978) sowie Ortmann (1987) dar, bei denen die Inhaftierten zufällig der Experimentalbedingung Sozialtherapie zugelost wurden. Dies konnte vor allem dadurch gerechtfertigt werden, dass zu diesem Zeitpunkt nur geringe Behandlungskapazitäten zur Verfügung standen und eine Randomisierung demzufolge keine evaluationsbedingte Benachteiligung der Kontrollgruppe bedeutete.

Der Einfluss der Designqualität auf die Studienergebnisse konnte mittels Metaanalysen für die Behandlungsforschung allgemein (Wilson & Lipsey, 2001) und für kriminalpräventive Studien im Besonderen (Weisburd, Lum & Petrosino, 2001) aufgezeigt werden. Je schwächer das Studiendesign, desto wahrscheinlicher ist ein Effekt zugunsten des Treatments und desto unwahrscheinlicher ist die Aufdeckung negativer Behandlungseffekte. Selbst zwischen experimentellen und quasiexperimentellen Untersuchungen, den beiden höchsten Stufen der Scientific Methods Scale, konnte diese Tendenz festgestellt werden, was die Bedeutung methodisch

hochwertiger Studien für die Wirksamkeitsbeurteilung von Behandlungsmaßnahmen unterstreicht.

Auch Randomisierung bietet jedoch keine Gewähr für uneingeschränkte Validität des Designs. Validitätsgefährdungen quasiexperimenteller und experimenteller Untersuchungen werden unterschieden in solche der statistischen Validität, der internen Validität, der Konstruktvalidität und der externen Validität (Cook & Campbell, 1979; Shadish, Cook & Campbell, 2002). Statistische Validität bezieht sich auf die Feststellbarkeit einer Kovariation zwischen zwei Variablen, interne Validität auf die Interpretierbarkeit einer Kovariation als kausal. Konstruktvalidität liegt vor, wenn durch die Operationalisierung tatsächlich das erfasst wird, was theoretisch erfasst werden soll und externe Validität besteht in der Generalisierbarkeit der Ergebnisse. Ohne auf alle denkbaren Validitätsgefährdungen einzugehen, werden im Folgenden einige von Shadish, Cook und Campbell genannte Aspekte kurz erwähnt, die für die Behandlungsforschung im Strafvollzug von Bedeutung sein können (vgl. Kury, 1983c).

Gefährdungen der Validität des statistischen Schlusses liegen häufig in zu geringer Teststärke, die aus nicht hinreichend großen Stichproben resultiert und zu dem fälschlichen Ergebnis führen kann, eine untersuchte Maßnahme hätte keinen Effekt (vgl. Lösel Köferl & Weber, 1987). Geringe Stichprobengrößen werden erreicht bei Evaluationen sehr spezifischer Angebote und insbesondere, wenn nicht ein ausreichender Zeitraum für die Untersuchung (z.B. für mehrere Behandlungsdurchgänge) einer Maßnahme zur Verfügung steht. Zur Überschätzung eines Behandlungseffekts kann das sogenannte „Signifikanzfischen“ führen, wenn multiple Vergleiche beispielsweise für eine Reihe von Effektivitätskriterien durchgeführt werden und in der Folge der Alphafehler kumuliert. Die zur Vermeidung dieses Fehlers häufig vorgeschlagene Alpha-Adjustierung (z.B. nach Bonferroni) behebt zwar das Problem eines irrtümlich signifikanten Ergebnisses, jedoch zuungunsten der bereits angesprochenen Teststärke und damit des Betafehlers. Zu einer Unterschätzung des Behandlungseffekts führen unreliable Maße sowie Decken- und Bodeneffekte bei der Erhebung der Effektkriterien sowie geringe Reliabilität des Treatments, wenn die untersuchte Maßnahme nicht hinreichend standardisiert implementiert wurde. Letzteres kann insbesondere der Fall sein, wenn Maßnahmen von unterschiedlichen Personen durchgeführt werden, das Personal unzureichend geschult ist, organisatorische Mängel vorliegen (Lösel & Bender, 1997) oder wenn Maßnahmen über lange Zeiträume hinweg untersucht werden und zwischenzeitlich Modifizierungen der Behandlungsform vorgenommen werden.

Gefährdungen der internen Validität liegen vor, wenn ein anderer Grund als die untersuchte Behandlungsmaßnahme für das Zustandekommen oder Nichtzustandekommen eines Ef-

fekts verantwortlich sein kann.⁷ Einflüsse auf die Behandlungsergebnisse können etwa von zwischenzeitlichem Geschehen oder Reifung zwischen Vor- und Nachtest ausgehen. Zwischenzeitliches Geschehen kann beispielsweise in der Veränderung von Haftbedingungen bestehen (z.B. dem Anstaltsklima, dem liberaleren oder restriktiveren Umgang mit Lockerungen) und Reifung kann sowohl persönliche Entwicklung als auch möglicherweise negative Hafteffekte (Prisonisierung bzw. Deprivation; Hosser, 2008; Müller-Marsell, 2004) beinhalten. Kontrolliert werden diese beiden Einflüsse zumeist, indem eine Vergleichsgruppe in derselben Anstalt rekrutiert wird, so dass beide Gruppen denselben Einflüssen ausgesetzt sind. Auch bei Verwendung eines Kontrollgruppendesigns ist die interne Validität jedoch nicht immer gesichert. Eines der größten Probleme der Behandlungsforschung im Strafvollzug liegt in Selektionseffekten, die nur bei randomisierter Behandlungszuweisung vollständig ausgeschlossen werden können. Geschieht dies nicht, so besteht die Gefahr, dass bereits vor der Behandlung bestehende Unterschiede zwischen Teilnehmern und Nichtteilnehmern einer Maßnahme als Behandlungseffekte fehlinterpretiert werden. Beispielsweise könnten Insassen mit besonders günstiger Prognose für eine Behandlung ausgewählt und die festgestellten positiven Veränderungen der Maßnahme zugeschrieben werden, obwohl die Veränderung auch unabhängig von der Behandlung eingetreten wäre. Eine weitere Problematik liegt im Stichprobenausfall, wenn dieser systematisch auftritt. Häufig sind gerade Behandlungsabbrecher nicht mehr zur Teilnahme an einer Nacherhebung zu motivieren, was dazu führen kann, dass der Behandlungseffekt überschätzt wird, da Abbrecher sich oft im Verlauf der Behandlung als behandlungsresistent erweisen.⁸

Gefährdungen der Konstruktvalidität stehen im Zusammenhang mit der Wahl geeigneter Effektivitätskriterien von Behandlung und liegen häufig – aber nicht ausschließlich – vor bei der Erhebung von Selbstangaben. So kann die Motivation der Probanden zur Behandlungsteilnahme die Testergebnisse beeinflussen, indem sich beispielsweise die Probanden im Prätest besonders behandlungsbedürftig darstellen, um in die angebotene Maßnahme zu gelangen. Diese Motivation würde für die Kontrollprobanden im Posttest anhalten während sie für die bereits Behandelten wegfällt, was einen fälschlichen Behandlungseffekt erzeugen würde.

⁷ Trochim (2001) unterscheidet Gefährdungen der internen Validität in Single-Group, Multi-Group und Social Interaction Threats. Während bei Eingruppenuntersuchungen Effekte wie zwischenzeitliches Geschehen, Reifung, Testung oder Stichprobenausfall die interne Validität gefährden, stellt bei Mehrgruppenuntersuchungen vor allem die Interaktion dieser Effekte mit der Selektion eine Validitätsgefährdung dar (Experimental- und Kontrollgruppe erfahren unterschiedliches zwischenzeitliches Geschehen, weisen unterschiedliche Reifung, etc. auf). Einschränkungen der Validität durch soziale Interaktion werden von Shadish, Cook und Campbell (2002) im Unterschied zu Trochim (2001) als Gefährdungen der Konstruktvalidität aufgefasst.

⁸ Im Kontext der strafvollzuglichen Behandlung nicht zu erwarten ist der in der allgemeinen psychotherapeutischen Behandlungsforschung mögliche Behandlungsabbruch seitens der Teilnehmer aufgrund einer Verbesserung der Symptomatik (Streiner, 2002), der bei Ausfall dieser Teilnehmer zu einer Unterschätzung der Behandlungseffektivität führen würde.

Versuchsleitereffekte sind insbesondere zu erwarten, wenn Behandlungsdurchführende auch Durchführende der Erhebung sind. Die Probanden könnten hier bemüht sein, den Erwartungen der Behandlungsdurchführenden im Sinne der Behandlungsziele zu entsprechen, oder aber diesen gezielt nicht zu entsprechen. Ähnlich wirkt das sogenannte „Hypothesis guessing“ bzw. der Hawthorne-Effekt, wenn Probanden Überlegungen über den Zweck der Untersuchung anstellen und sich als besonders „gute“ Versuchspersonen geben wollen. Gefährdungen der Konstruktvalidität unabhängig von der Methode der Erhebung liegen vor, wenn Elemente der Behandlung durch Kontakt der Untersuchungsteilnehmer auf die Kontrollgruppe übertragen werden, wenn also die Behandlung diffundiert. Dass aus der bloßen Kenntnis von Behandlungselementen jedoch ein Behandlungseffekt resultiert, ist im strafvollzuglichen Kontext weniger zu erwarten. Dagegen ist ein möglicher kompensatorischer Ausgleich von Behandlung nicht selten anzutreffen. Dieser liegt vor, wenn Nichtteilnehmer einer Maßnahme (z.B. des Anti-Aggressivitäts-Trainings) tendenziell häufiger einer anderen Maßnahme (z.B. dem sozialen Training) zugewiesen werden. Hier würde im Extremfall eine als isoliert angelegte Evaluation nur einer Maßnahme zu einer vergleichenden Evaluation beider Maßnahmen.⁹ Weitere sich nicht zwingend auf Selbstangaben beziehende reaktive Tendenzen der Probanden sind Demoralisierung oder Rivalität der Kontrollgruppe als Reaktion auf die Erfahrung, nicht in die Behandlungsgruppe aufgenommen worden zu werden, was entweder zu einer Über- oder Unterschätzung des Behandlungseffekts führen würde.

Festzuhalten bleibt, dass eine Vielzahl möglicher Validitätsgefährdungen die Aussagekraft empirischer Wirksamkeitsuntersuchungen der Straftäterbehandlung einschränken können. Durch die von Vertretern der evidenzbasierten Kriminalprävention geforderte Realisierung experimenteller Designs allein kann diesem Umstand nicht begegnet werden. Kritiker des Verfahrens von Sherman et al. (1997) weisen zudem auf den erhöhten Typ II Fehler (der Klassifikation eines wirksamen Programms als unwirksam; Hope, 2005), sowie auf die alleinige Fokussierung auf die Wirkungsforschung hin. Nicht allein ob eine Maßnahme wirkt, sondern auch warum und unter welchen Umständen seien wichtige Fragen der Evaluationsforschung, die beim dargestellten Ansatz unberücksichtigt blieben (z.B. Leeuw, 2005).¹⁰

⁹ Vgl. die Evaluation des Anti-Aggressivitäts-Trainings von Ohlemacher et al. (2001) mit dem Titel „Nicht besser, aber auch nicht schlechter.“ Die Autoren schlussfolgern aus den nichtsignifikanten Ergebnissen ihrer Studie, dass das Training eine der allgemeinen Behandlung der Jugendanstalt vergleichbare Wirksamkeit aufweise. Von Bedeutung ist in diesem Zusammenhang auch, dass bei einer vergleichenden Evaluation höhere Effektstärken bzw. größere Stichprobenumfänge für die Feststellung signifikanter Ergebnisse benötigt werden als bei einer isolierten Evaluation (Hager, Leichsenring & Schiffler, 1999).

¹⁰ Illustriert wird dieser Umstand in der Debatte zwischen Hans Zeisel und Peter Rossi in den frühen 80er Jahren (Rossi, Berk & Lenihan, 1980, 1982; Zeisel, 1982a, 1982b). Obwohl sich in einem randomisierten Experiment zur Vergabe finanzieller Hilfen an Haftentlassene die Experimentalgruppe als ebenso rückfällig wie die Kontrollgruppe erwies, resümierte die Forschungsgruppe um Rossi einen positiven Effekt der Intervention. Begrün-

So bleibt insbesondere bei fehlendem Wirkungsnachweis offen, ob die dem Programm zugrunde liegenden theoretischen Annahmen falsch waren, die Maßnahme zu gering dosiert oder fehlerhaft implementiert wurde. Alternativ werden Evaluationsansätze wie „theory-driven“ oder „realistic evaluation“ (Chen & Rossi, 1983; Pawson & Tilley, 1994) vorgeschlagen, die die Black-box-Problematik der reinen Wirkungsforschung aufheben. Theoriegeleitete Evaluation beginnt mit der Formulierung eines Wirkmodells und erhebt und überprüft auch solche Merkmale, die als zwischengelagerte Bedingungen der Programmwirksamkeit angenommen werden (Chen & Rossi, 1983). Hierdurch können Hinweise darauf gewonnen werden, an welcher Stelle des überprüften Modells theoretische Annahmen gegebenenfalls unzutreffend waren oder praktische Probleme der Programmumsetzung eine positive Programmwirksamkeit behindert haben. Auf diesen Punkt wird bei der Vorstellung verschiedener Effektivitätskriterien strafvollzoglicher Behandlung noch eingegangen.

Im Folgenden werden eigene Untersuchungen zur Behandlungswirksamkeit im Jugendstrafvollzug unter Berücksichtigung möglicher Einschränkungen der Validität vorgestellt. Besonderes Augenmerk wird in diesem Zusammenhang darauf gerichtet, ob trotz der fehlenden Realisierung randomisierter Untersuchungsdesigns praxisrelevante Befunde festgestellt werden können.

det wurde dies durch Befunde innerhalb der Experimentalgruppe: So waren unterstützte Personen zum einen seltener oder in geringerem Ausmaß erwerbstätig und zum anderen stand Erwerbstätigkeit in negativem Zusammenhang mit der Rückfälligkeit. Beides zusammen ergäbe einen Effekt der Interventionsmaßnahme auf Rückfälligkeit, der nur durch einen zusätzlichen, nicht beobachteten Wirkungspfad ausgehend von der Intervention auf Rückfälligkeit ausgeglichen werden könne. Das Argument lautete, dass wenn man durch andere Modalitäten der Mittelvergabe erreichen könne, dass finanziell unterstützte Haftentlassene im selben Umfang Arbeit aufnehmen wie nicht unterstützte Haftentlassene, müsse der postulierte positive Effekt des Programms zum Tragen kommen (Rossi, Berk & Lenihan, 1980).

Zeisel kritisierte an diesem Vorgehen, dass die Schlussfolgerungen nicht auf Grundlage des experimentellen Untersuchungsdesigns getroffen würden, sondern auf Basis korrelativer Befunde. Neben der geringeren methodischen Güte dieser Analyse im Vergleich zum Experiment wies er zudem darauf hin, dass Rossi et al. nicht angeben könnten, wie genau eine andere Form der Mittelvergabe zu realisieren sei, damit sie das gewünschte Ziel erreiche (Zeisel, 1982a). Und selbst angenommen, eine solche Form der Mittelvergabe wäre bekannt, so müsse die mögliche Programmwirksamkeit erst durch die erneute Durchführung eines Experiments nachgewiesen werden (Zeisel, 1982b).

Vereinfacht gesagt zeigen sich hier unterschiedliche Auffassungen von Evaluationsforschung, bzw. der in ihrem Fokus liegenden kausalen Prozesse (Smith, 1990). So kann Evaluation als methodisch stringent geführte Wirksamkeitsüberprüfung im Sinne einer Outcome-Evaluation verstanden werden, ebenso jedoch im Sinne der Erforschung zugrundeliegender Mechanismen der Wirksamkeit. Es ist einerseits Zeisel damit rechtzugeben, dass Rossi et al. den empirischen Nachweis der Wirksamkeit der Maßnahme nicht erbringen konnten. Andererseits tragen aber die von Rossi et al. aufgedeckten positiven und negativen Wirkungspfade zum Verständnis der Wirkmechanismen der Intervention bei, sowie möglicherweise auch zu deren Fortentwicklung – und eben dies sind zentrale Aufgaben der Evaluationsforschung, die über die rein summative Betrachtung der Programmwirksamkeit hinausgehen.

3.2.1 Psychosoziale Behandlung im Jugendstrafvollzug

Im Projekt „Entwicklungsfolgen der Jugendstrafe“ wurde der Zusammenhang der Teilnahme an verschiedenen psychosozialen Behandlungsmaßnahmen in Haft mit Risikofaktoren für delinquentes Verhalten und offiziell registrierter Rückfälligkeit untersucht (Bosold, Hosser & Lauterbach, 2007). Im Einzelnen handelt es sich dabei um soziales Training, Sozialtherapie, Suchtherapie, Suchtberatung, Anti-Gewalt-Training, sonstige Gruppentrainings und Einzel-fallhilfen (weitere Beratungsangebote und Entlassungsvorbereitungskurse wurden nicht be-rücksichtigt). An mindestens einer dieser Maßnahmen haben 55 % der Insassen teilgenommen (38 % der sachsen-anhaltinischen und 70 % der niedersächsischen Stichprobe). Behandelte und Unbehandelte unterscheiden sich zwar tendenziell, aber nicht signifikant hinsicht-lich ihrer Rückfälligkeit (siehe Abbildung 3). Als einziger Behandlungseffekt lässt sich im Prä-Post-Vergleich eine Verbesserung der moralischen Urteilsfähigkeit feststellen, keine Unterschiede finden sich dagegen bei der Aggressivität, den Neutralisierungstendenzen und verschiedenen Aspekten der Prisonisierung.

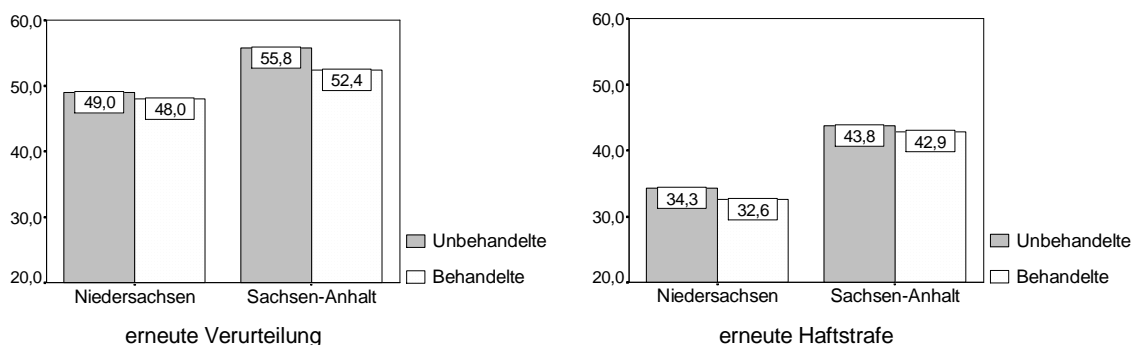


Abbildung 3: Quote der erneuten Verurteilung und erneuten Haftstrafe innerhalb von 12 Monaten in Abhängigkeit von Behandlung

In der vorliegenden Untersuchung wurde ein quasiexperimentelles Kontrollgruppendesign realisiert, wobei die Experimentalbedingung Behandlung retrospektiv aus den Gefangenepersonalakten erfasst wurde. Hinsichtlich der von Cook und Campbell (1979) genannten Gefährdungen der Validität von Untersuchungsdesigns liegen hierin sowohl Vor- als auch Nachteile begründet. Von Vorteil ist, dass den Befragten die Forschungsintention der Behandlungsevaluation verborgen bleibt. Gefährdungen der internen Validität wie Rivalität oder Motivationsverlust der Kontrollgruppe sind demzufolge nicht im selben Ausmaß zu erwarten wie bei der Evaluation einer konkreten Behandlungsmaßnahme. Nachteile der Erfassung der Behandlungsteilnahme aus den Personalakten bestehen darin, dass die Maßnahmen nur relativ grob kategorisiert sind und selbst bei Betrachtung einer einzelnen Maßnahme nicht sicherge-

stellt ist, dass diese nicht in unterschiedlichen Anstalten oder bei unterschiedlichen Behandlungsdurchführenden in ihrer Zielsetzung oder Implementation variiert (geringe Reliabilität des Treatments). Das weitaus größte methodische Problem resultiert aber aus der Behandlungszuweisung, die an individuellen Bedürfnissen und Fähigkeiten des Straftäters ausgerichtet ist. Dies bedeutet, dass Unbehandelte oder anderen Maßnahmen zugewiesene Inhaftierte nicht dieselben Voraussetzungen aufweisen (Selektionseffekte) und damit nur eingeschränkt als Vergleichsgruppe geeignet sind.

In der Folge lassen sich aus der vorgestellten Untersuchung streng genommen keine gesicherten Aussagen über die Behandlungs(un)wirksamkeit ableiten, da im Unklaren bleibt, worin die zuweisungsentscheidenden Merkmale bestehen und Interaktionen der Zuweisung mit anderen Störfaktoren nicht ausgeschlossen sind. Die nicht nachweisbaren Behandlungseffekte könnten darauf zurückzuführen sein, dass in der Behandlungskonferenz Insassen selektiert wurden, die sich zwar in ihren Vortestwerten des Evaluationskriteriums (Aggressivität, Neutralisierung, moralische Urteilsfähigkeit) nicht von den Unbehandelten unterscheiden, die aber aufgrund anderer nicht kontrollierter Merkmale (z.B. kognitive Fähigkeiten oder Behandlungsmotivation) eine schlechtere Prognose aufweisen. In der Praxis werden jedoch eher solche Personen einer Behandlung zugeführt, bei denen neben der Therapiebedürftigkeit auch Therapiefähigkeit und Therapiemotivation gegeben sind (Bosold, 2008; Dahle, 1997), so dass eher von einer Positiv- als von einer Negativselektion auszugehen ist. Eine andere mögliche Erklärung für die nicht nachweisbare Behandlungseffektivität ist der konservative Umgang mit Behandlungsabbrechern, die bei der Auswertung in der Gruppe der Behandelten verblieben. Dies entspricht dem in der medizinischen Forschung üblichen Intention-to-treat-Ansatz (Hollis & Campbell, 1999) welcher verhindert, den Selektionsprozess des Verbleibs in der Behandlungsmaßnahme fälschlicherweise als Behandlungseffekt zu interpretieren. Da Behandlungsabbrecher mitunter höhere Rückfälligkeit als Unbehandelte aufweisen (z.B. Schmucker & Lösel, 2007) hätte die alternative Auswertungsstrategie des separaten Ausweisens der Behandlungsabbrecher (vgl. Dünkel, 1980; Ortmann, 2002) möglicherweise zu größeren Behandlungseffekten geführt.

Trotz des fehlenden Nachweises von Behandlungswirksamkeit leistet die Untersuchung einen bedeutsamen Erkenntnisfortschritt. So kann festgehalten werden, dass der Jugendstrafvollzug global betrachtet keine Veränderung wichtiger Risikomerkmale delinquenten Verhaltens (Aggressivität und Neutralisierung der Straftat) bewirkt. Erinnerung sei in diesem Zusammenhang an die Bezeichnung eines Gefängnisaufenthalts als „behavioral deep freeze“ (Zamble & Porporino, 1990), die impliziert, dass Inhaftierte mit denselben risikobelasteten

Verhaltensdispositionen aus dem Vollzug entlassen werden, mit denen sie inhaftiert wurden. Als ein weiterer Befund scheinen sich negative Hafteffekte wie Autonomiebegrenzung und Bindung an die Mitinsassen im Haftverlauf noch zu verstärken. Hoffnungsvoll stimmen dagegen die Verbesserung des moralischen Urteilsvermögens und differentielle Effekte einzelner Behandlungsmaßnahmen, die auf eine Wirksamkeit beispielsweise des Anti-Gewalt-Trainings hinsichtlich der Verringerung von Aggressivität und der Sozialtherapie hinsichtlich des Ausgleichs von Prisonisierungseffekten deuten – ohne diese allerdings im methodisch strengen Sinn zu belegen.

3.2.2 Sozialtherapeutische Behandlung von jungen Sexualstraftätern

Zur Untersuchung der Behandlungswirksamkeit der neu aufgebauten sozialtherapeutischen Abteilung für Sexualstraftäter der Jugendanstalt Hameln wurden standardisierte Befragungen der Behandlungsteilnehmer durchgeführt (Hosser, Bosold & Lauterbach, 2006). Im Prä-Post-Vergleich zeigte sich eine Verringerung der feindseligen sexuellen Überzeugungen, der Verharmlosung von Kindesmissbrauch sowie der situationsbezogenen Aggressivität. Entgegen der Behandlungsziele konnte eine Abnahme der Flexibilität der Zielanpassung festgestellt werden. Weiterhin konnten keine Veränderungen in anderen wichtigen Effektivitätskriterien der Behandlung, darunter Verharmlosung von Vergewaltigung, Empathie und Opferempathie, Impulsivität und Neutralisierung festgestellt werden.

Ein besonderes Problem der vorliegenden Evaluationsstudie besteht darin, dass innerhalb der Jugendanstalt – in Anlehnung an die gesetzliche Regelung für den Erwachsenenstrafvollzug (vgl. Dünkel, 2005; Egg, 2000) – alle Insassen, die wegen Vergewaltigung oder Kindesmissbrauch verurteilt wurden, der Abteilung zugeführt und zunächst in die Motivationsgruppe aufgenommen wurden. In die Behandlungsgruppe wurde zu individuell durch die Therapeuten bestimmten Zeitpunkt nur ein Teil der Sexualdelinquenten übernommen. Es ließ sich folglich keine adäquate Kontrollgruppe realisieren, da nicht in die Behandlungsgruppe übernommene Inhaftierte systematische Unterschiede bezogen auf Behandlungsfähigkeit oder -motivation zu den Behandelten aufgewiesen hätten. Eine zunächst geplante anstaltsübergreifende Lösung scheiterte an zu hohem organisatorischem Aufwand. In der Konsequenz können die festgestellten Veränderungen nicht mit Sicherheit als Behandlungseffekte interpretiert werden, da Reifungseffekte, zwischenzeitliches Geschehen sowie Selektionsprozesse durch die relativ hohe Rückverlegungsquote Einfluss auf die Nachbefragungsergebnisse genommen haben können. Ebenfalls problematisch ist, dass aus Gründen knapper Ressourcen die Erhebungen durch das Behandlungsteam durchgeführt wurden, so dass Antworttendenzen wahrscheinli-

cher sind als bei einer unabhängigen Befragung. Eine Einschränkung der Untersuchung die sich eher im Sinne der potentiellen Unterschätzung des Behandlungseffekts auswirkt liegt in der Stichprobengröße und der damit verbundenen geringen Teststärke, wodurch sich erst Veränderungen mit relativ großer Effektstärke statistisch abgesichert feststellen lassen.

Obwohl das Evaluationsdesign unter Validitätsaspekten als nicht optimal einzustufen ist, können Ergebnisse von hoher praktischer Relevanz festgestellt werden. Zum einen scheint die untersuchte Maßnahme eine deliktspezifische Wirksamkeit aufzuweisen, da zwar die Verharmlosung von Kindesmissbrauch, nicht aber die von Vergewaltigung zurückging. Als ein weiterer Befund bleibt festzuhalten, dass sich gerade für die im Bezug auf Sexualdelinquenz als besonders wichtig erachteten Aspekte Opferempathie und Neutralisierung der Tat keine Verbesserungen erreichen ließen. Konsequenzen sollten also darin bestehen, sich verstärkt der Tätergruppe der Vergewaltiger zu widmen und ihnen die Tatfolgen zu verdeutlichen sowie Neutralisierungstendenzen entgegenzuwirken und Empathiefähigkeit zu trainieren. Die vorgestellte Untersuchung hat somit insgesamt eher begleitenden Charakter im Sinne der Behandlungsoptimierung und ist aufgrund der erwähnten Einschränkungen nicht als abschließender Wirksamkeitsnachweis zu verstehen.

3.2.3 Evaluation eines Anti-Aggressivitäts-Trainings

Das LoGo-Training in der Jugendanstalt Hameln stellt eine Weiterentwicklung des ursprünglich in Hameln entwickelten Anti-Aggressivitäts-Trainings dar. Bei der Evaluation des Trainings zeigte sich im Vergleich zu einer Wartekontrollgruppe in Selbstangaben eine Verringerung verschiedener Aggressivitäts-Maße (siehe Abbildung 3) sowie von Neutralisierungstendenzen (Bosold & Lauterbach, 2010). Die durch das Training ebenfalls angestrebte Ressourcenstärkung hinsichtlich Empathiefähigkeit, Selbstwert und akkommodativer Bewältigungsstrategien ließ sich dagegen nicht nachweisen. Weiterhin fanden sich Hinweise darauf, dass erreichte Behandlungseffekte mit längerem Beobachtungszeitraum abnahmen und hinsichtlich offiziell registrierter Gewaltrückfälligkeit innerhalb eines Jahres konnten keine Unterschiede zwischen Behandelten und Unbehandelten festgestellt werden.

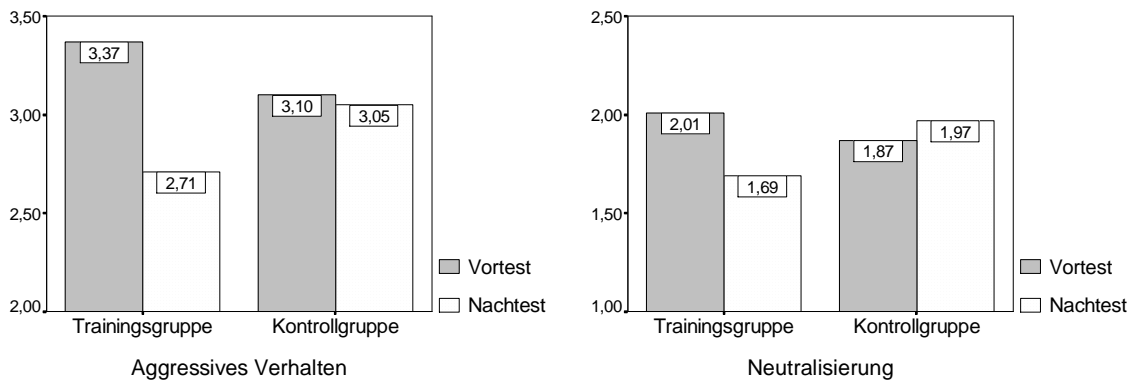


Abbildung 4: Veränderung von Aggressivität und Neutralisierung bei Trainingsteilnehmern im Vergleich zur Kontrollgruppe

Anders als in der vorangegangenen Studie konnte bei der Evaluation des LoGo-Trainings auf eine Kontrollgruppe unbehandelter Gewaltstraftäter zurückgegriffen werden, die nach Behandlungsplan für das Training vorgesehen waren, aber aufgrund zu geringer Kapazität nicht aufgenommen wurden. Somit sind Reifungseffekte und Einflüsse zwischenzeitlichen Geschehens weitestgehend auszuschließen, da diese ebenso die Kontrollgruppe betroffen hätten. Nicht grundsätzlich ausgeschlossen sind dagegen Diffusion oder Imitation des Treatments, kompensatorischer Ausgleich der Behandlung, kompensatorische Rivalität der Probanden, Motivationsverlust der Kontrollprobanden und Einfluss des Stichprobenausfalls.

Auch wenn sich unter den Inhaftierten Informationen über einzelne Behandlungselemente wie etwa Anti-Blamier-Übungen oder den „heißen Stuhl“ verbreiten, ist hiervon keine Behandlungswirksamkeit in der Kontrollgruppe zu erwarten, so dass von einem Diffundieren des Treatments nicht ausgegangen werden kann. Ebenso ist eine Rivalität der Kontrollprobanden aufgrund der eher behandlungskritischen Einstellung der Inhaftierten unwahrscheinlich. Dagegen kann ein kompensatorischer Ausgleich von Behandlung nicht völlig ausgeschlossen werden, da es plausibel erscheint, dass Nichtteilnehmer der Maßnahme verstärkt in andere Maßnahmen wie etwa das soziale Training vermittelt wurden und die Teilnahme an anderen Behandlungsmaßnahmen in der vorliegenden Studie nicht kontrolliert wurde. Allerdings setzen andere Maßnahmen nicht an denselben kriminogenen Risikofaktoren an, so dass die Maßnahmen nicht als äquivalent hinsichtlich ihrer Nahziele anzusehen sind. Gemeinsam ist den genannten Validitätsgefährdungen zudem, dass sie eher konservativ in Richtung der Unterschätzung eines angenommenen Behandlungseffekts des LoGo-Trainings wirken würden.

Zu einer möglichen Überschätzung des Behandlungseffekts könnten die Demoralisierung der Kontrollprobanden sowie der Stichprobenausfall beitragen. Gegen eine mögliche Demoralisierung spricht, dass sich keine deutlichen Unterschiede zwischen den Vor- und Nachtest-

werten der Kontrollprobanden fanden. Allein der Stichprobenausfall stellt in der vorliegenden Untersuchung ein Problem dar, da Abbrecher des Trainings oft nicht zur Teilnahme an der Nachbefragung zu motivieren waren. Analysen, bei denen im Sinne eines Intention-to-treat-Ansatzes fehlende Werte bei der Nachbefragung durch die Vortestwerte ersetzt wurden (Last-observation-carried-forward-Methode; Shao & Zhong, 2003) führten allerdings nicht zu wesentlichen Veränderungen der Befunde.

Zusammenfassend betrachtet liefert die Studie unter den vorgestellten Untersuchungen die am besten abgesicherten Effekte hinsichtlich behandlungsspezifischer Nahziele. Die Ergebnisse decken sich zum Teil mit den bereits berichteten Befunden von Bosold, Hosser und Lauterbach (2007), wo für die weiter gefasste Kategorie der Anti-Gewalt-Trainings eine tendenzielle Verringerung der Aggressivität festgestellt werden konnte. Unklar bleibt allerdings, warum sich Effekte zwar in der Reduzierung der Gewaltneigung, nicht jedoch in der Legalbewährung zeigen: Bislang fanden sich keine Unterschiede zwischen Behandelten und Unbehandelten hinsichtlich offiziell registrierter Gewaltdelikte nach Haftentlassung, wobei die Überprüfung eines längeren Rückfallzeitraums noch aussteht. Eine mögliche Erklärung für das Ausbleiben von Effekten in der Legalbewährung liefert die zeitlich begrenzte Stabilität der erreichten Einstellungs- und Verhaltensänderungen, die eine Auffrischung der Trainingsinhalte in bestimmten Zeitintervallen ratsam erscheinen lassen. Ein weiterer Grund könnte darin liegen, dass sich eine Ressourcenstärkung bei den Trainingsteilnehmern nicht nachweisen ließ – eine zentrale Annahme der Gewaltprävention ist jedoch, dass neben dem Eindämmen von Risikofaktoren personale Ressourcen sowie Schutzfaktoren aufgebaut werden müssen (vgl. Gollwitzer, 2007). Aus methodischer Perspektive betrachtet bringt außerdem Rückfälligkeit als Kriterium von Behandlungswirksamkeit besondere Einschränkungen mit sich, wie im folgenden Abschnitt aufgezeigt wird.

3.3 Effektkriterien der Behandlungsforschung

Das übergeordnete Ziel des Strafvollzugs ist die Rückfallprävention (Behnke, 2004), aus diesem Grund ist das Legalverhalten aus dem Strafvollzug entlassener Personen ein naheliegender und weit verbreiteter Indikator für die Effektivität von Sanktionen im Allgemeinen und für strafvollzugliche Behandlung im Besonderen (Feltus, 1991; Kury, 1986). Unterschiedliche Rückfalldefinitionen beziehen sich auf das Vorliegen einer Rückfalltat im Hell- oder Dunkelfeld, verschiedene Schweregrade der Sanktion sowie die Art des Rückfalldelikts (Egg, 2007, Wirth, 1996).

Ein Hauptgrund der weiten Verbreitung des Kriteriums offiziell registrierter Rückfälligkeit ist neben der scheinbaren Objektivität die leichte Verfügbarkeit durch Auswertung des Bundeszentral- und Erziehungsregisters (Lösel, Köferl & Weber, 1987). Vergleicht man einzelne Rückfallstudien, so finden sich jedoch eine Vielzahl unterschiedlicher Definitionen und Handhabungen (vgl. Heinz & Jehle, 2004). So werden mitunter nur Rückfälle einer bestimmten Deliktgruppe wie etwa Gewalt- oder Sexualdelikte (z.B. Egg, 2004; Harrendorf, 2006) oder nur einschlägiger Rückfall (z.B. eine erneute Sexualstraftat eines verurteilten Sexualstraftäters) gewertet. Eine andere Unterscheidung betrifft die Schwere der Sanktion, bei der beispielsweise zwischen Geldstrafen und bedingten sowie unbedingten Haftstrafen unterschieden werden kann. Weiterhin variiert häufig der Beobachtungszeitraum zwischen verschiedenen Untersuchungen, was wiederum zu unterschiedlichen Rückfallhäufigkeiten führt (Wirth, 1996).¹¹

Neben Kriminalität und Rückfälligkeit im Hellfeld ist die Kriminalität im Dunkelfeld ein Problem, das die Validität eines allein an der offiziell registrierten Kriminalität ausgerichteten Rückfallkriteriums einschränkt (Brody, 1976; Kury, 1983c). In einzelnen Deliktbereichen ist davon auszugehen, dass die Mehrzahl der verübten Straftaten ohne strafrechtliche Konsequenzen für den Täter bleibt, weil die Tat nicht angezeigt wurde oder der Täter nicht zu ermitteln war (Heinz, 2006; Sack, 1993). In der Folge wird mit offiziell registrierter Rückfälligkeit nur ein Teil des tatsächlichen Legalverhaltens einer Untersuchungsstichprobe abgebildet. Prinzipiell bieten hier Dunkelfeldbefragungen einen Ausweg, bei denen in Form von Selbstangaben kriminelles Verhalten erfasst wird; dies geschieht jedoch in der Regel anhand leicht zugänglicher Stichproben wie etwa Schülern (Kerner, 1991). Im Falle selbstberichteter Delinquenz als (Miss-)Erfolgskriterium von Behandlung muss jedoch ein klar definierter Personenkreis von entlassenen Straftätern weiter untersucht werden, was in der Praxis hohen organisatorischen Aufwand und fast unausweichlich Stichprobenausfälle mit sich bringt. Zudem gelten die bei der Erhebung von Selbstangaben üblichen Einschränkungen aufgrund von Antworttendenzen, insbesondere bei nicht ausreichend gesicherter Anonymität der Befragung.

Kritiker einer alleinigen Berücksichtigung der Legalbewährung als Erfolgskriterium von Behandlung bemängeln die oft fehlende theoretische Verknüpfung von vollzoglicher Intervention und erneuter Straffälligkeit (z.B. Wirth, 1996). Für eine über reine Wirksamkeitsfeststellung hinausgehende Evaluation von Interventionsmaßnahmen bedarf es eines theoretischen Rahmens, der das Verhältnis von Behandlungselementen und direkten sowie indirekten Behandlungszielen beschreibt (Lösel, Köferl & Weber, 1987). Im allgemeineren Kontext der

¹¹ Wird Rückfall anhand von Bundeszentralregisterdaten erhoben, so müssen die dortigen Lösungsfristen berücksichtigt werden (Jehle, 2007).

Evaluationsforschung entspricht dies der bereits angesprochenen Unterscheidung von reiner Outcome-Evaluation und theoriegeleiteter Evaluation (Chen & Rossi, 1983; Weiss, 1997).

Anstatt mit dem Legalbewährungskriterium allein am Ende der Wirkungskette anzusetzen, werden in vielen Untersuchungen für die jeweiligen Maßnahmen spezifische Ziele von Behandlung, wie z.B. die Veränderung kriminogener Risikofaktoren überprüft. Solche Ziele sind beispielsweise bei der Gewalttäterbehandlung die Senkung der Gewaltneigung und Impulsivität, bei Drogentherapien die Aufgabe oder Verminderung des Substanzkonsums und für Bildungsmaßnahmen das Erlangen entsprechender Qualifikationen. Vorteile dieser behandlungsspezifischen Kriterien liegen neben ihrer besseren inhaltlichen Verknüpfung mit der Maßnahme in der leichteren Erfassbarkeit, da sie zumeist noch während der Haftstrafe und üblicherweise im direkten Anschluss an die Behandlungsmaßnahme erhoben werden können. Allerdings stellt sich hier – mit Ausnahme der formalen Qualifikationen – das Problem der Stabilität der Effekte. So ist nicht gesichert, dass bei Veränderung z.B. von Verhaltensdispositionen wie Aggressivität die Übertragung vom Haftkontext in das Leben in Freiheit gelingt. Weiterhin ist das Verhältnis von Erreichung des Nahziels einer Behandlung und des Fernziels der Legalbewährung häufig unklar. So ist nicht immer gesichert, dass mit der Reduzierung eines Risikofaktors durch Behandlung das allgemeine Rückfallrisiko gesenkt wird, da möglicherweise andere Risikofaktoren unbehandelt geblieben sind (zum Problem der Komorbidität siehe Lösel & Bender, 1997).

Im Zuge der Evaluation von Behandlungsprogrammen für Sexualstraftäter wurden Auswertungsmethoden diskutiert, die verschiedene Effektivitätskriterien von Behandlung miteinander verknüpfen (Hanson, 1997; Marques, 1999). Es herrscht weitgehender Konsens, dass auf das Legalbewährungskriterium nicht verzichtet werden sollte, auch wenn Befürchtungen geäußert werden, dieses sei für die Behandlungsevaluation in vielen Fällen zu insensitiv (Barbaree, 1997).¹² Ein interessanter Ansatz besteht in der schrittweisen Überprüfung von Behandlungswirksamkeit, bei der zunächst an den direkten Behandlungszielen orientierte Within-treatment-Veränderungen kriminogener Risikofaktoren als Effektkriterien verwendet werden, die in weiteren Analysen als Prädiktoren der Legalbewährung überprüft werden (Harkins & Beech, 2006). Dieses Vorgehen erlaubt einerseits Aussagen über eine Maßnahmenwirksamkeit, die weniger stark von Umwelteinflüssen und Störfaktoren beeinträchtigt

¹² Mittels des Legalbewährungskriteriums lassen sich in Abhängigkeit der Basisrate der Rückfälligkeit und der Stichprobengröße nur relativ große Behandlungseffekte nachweisen (Barbaree, 1997). Zum Nachweis geringerer Effekte sind größere Stichproben oder zur Erhöhung der Basisrate insbesondere bei delikt spezifischem Rückfall längere Beobachtungszeiträume notwendig. Die Stichprobengröße ist jedoch in der Regel an behandlungspraktische Gegebenheiten gekoppelt und bei Berücksichtigung längerer Beobachtungszeiträume ist mit wachsender Bedeutung kaum kontrollierbarer Umwelteinflüsse zu rechnen, die der Stabilität der Behandlungseffekte zuwiderlaufen (Lösel & Bender, 1997).

werden und andererseits im Sinne einer theoriegeleiteten Evaluation anhand der jeweiligen Nahziele von Behandlung die Identifikation von Behandlungselementen, die sich möglicherweise als nicht funktional erwiesen haben.

Ein weiteres Glied in der Wirkungskette Erreichung behandlungsspezifischer Nahziele und des Fernziels Rückfallvermeidung stellt die Sozialbewährung oder Resozialisierung dar (Cornel, 1995), von der angenommen wird, dass sie der Legalbewährung kausal vorausgeht (Ortmann, 2002). Insbesondere für jugendliche Straftäter kann auch die Bewältigung von Entwicklungsaufgaben (z.B. Havighurst, 1972) als ein (Re-)Sozialisierungsziel angesehen werden (Greve, 1998). Als relativ leicht zu erfassende Merkmale von Resozialisierung bzw. altersgemäßer Entwicklung können dabei im Einzelnen die Wohn- und Erwerbssituation, familiäre und partnerschaftliche Bindungen, Drogenabstinenz und soziale Einbindung in der Freizeit betrachtet werden. Als Effektivitätskriterien von Behandlung werden solche Indikatoren allerdings selten herangezogen, obwohl gerade die soziale und berufliche Integration Ziel vieler Interventionsmaßnahmen ist. Zwar ist ähnlich wie bei der Erfassung selbstberichteter Delinquenz die Erhebung verglichen mit anderen Kriterien vergleichsweise aufwändig, im Zuge einer ohnehin notwendigen Verzahnung von strafvollzuglicher Behandlung, Entlassungsvorbereitung und Nachbetreuung (gefordert z.B. von Bertram, 2004; Wirth, 2006a) erscheint es jedoch zumindest denkbar, Merkmale der Nachentlassungssituation von Straftätern systematisch zu erheben und als Effektkriterien von Behandlung heranzuziehen.

Zusammenfassend betrachtet ist Rückfall als alleiniges Erfolgskriterium von Behandlung im Strafvollzug abzulehnen, da Probleme der Definition und Erfassung, der theoretischen Fundierung sowie der Sensitivität bestehen. Viele Studien berücksichtigen stattdessen spezifische Nahziele von Behandlung als Erfolgskriterien, wie z.B. Veränderungen kriminogener Risikofaktoren, die jedoch für sich allein genommen ebenfalls keine abschließenden Aussagen über Behandlungswirksamkeit erlauben. Erst eine theoretisch gut fundierte Kombination sich auf Nah- und Fernziele von Behandlung beziehender Effektivitätskriterien erlaubt eine abschließende Wirksamkeitsbeurteilung (Lösel & Bender, 1997). Als eine in dieser Arbeit verfolgte These sollten dabei auch Merkmale der sozialen Integration im Nachentlassungszeitraum berücksichtigt werden, was derzeit noch in wenigen Untersuchungen geschieht (als eine Ausnahme z.B. Boxberg & Bosold, 2009). Im Folgenden werden deshalb Untersuchungen dargestellt die sich mit der Nachentlassungssituation junger Straftäter beschäftigen, um Hinweise auf aus diesem Kontext stammende Effektkriterien von Behandlung im Jugendstrafvollzug zu gewinnen.

3.3.1 Entlassungsvorbereitung und Nachentlassungssituation

Im weiteren Sinn soll bereits die Vollzugsplanung gemäß des Resozialisierungsziels auf das zukünftige Leben in Freiheit vorbereiten, im engeren Sinn besteht die Entlassungsvorbereitung in beratenden und unterstützenden Maßnahmen, die den Übergang vom Vollzug in das Leben außerhalb der Strafanstalt erleichtern sollen. Die Angebote werden sowohl als Einzelberatungen als auch in Form von Entlassungsvorbereitungskursen durchgeführt und richten sich an Inhaftierte, deren Entlassungszeitpunkt kurz bevorsteht.

Anhand der Befragungsdaten aus dem Projekt „Entwicklungsfolgen der Jugendstrafe“ werden Merkmale und Ergebnisse von Entlassungsvorbereitung im Jugendstrafvollzug näher beleuchtet (Hosser, Lauterbach & Höynck, 2007). Zunächst ist festzustellen, dass nur knapp 60 % der jungen Inhaftierten an einer entlassungsvorbereitenden Maßnahme im engeren Sinn teilgenommen haben. In dieser Maßnahme waren die am häufigsten behandelten Themen die Suche nach einem Arbeitsplatz, Hilfe bei Behördenangelegenheiten, Kontaktaufnahme zu Beratungsstellen und Angehörigen sowie die Wohnungssuche. Dass das Angebot möglicherweise zu abstrakt und nicht immer von konkretem Nutzen für die Betroffenen ist, zeigt sich an der Vermittlungsquote für Arbeits- und Ausbildungsverhältnisse. So konnte zwar in 36 % der Fälle eine Beschäftigung im Anschluss an die Inhaftierung in Aussicht gestellt werden aber nur die Minderheit von 7 % konnte mit der Entlassung direkt in ein Beschäftigungsverhältnis wechseln bzw. darin verbleiben, wenn sie dort bereits als Freigänger beschäftigt waren.

Die besondere Bedeutung entlassungsvorbereitender Hilfen zeigt sich in der schwierigen Lebenssituation der Befragten durchschnittlich ein dreiviertel Jahr nach Entlassung. Ein ernüchternder Befund ist, dass trotz schulischer und beruflicher Förderung in Haft sowie Unterstützung durch Entlassungsvorbereitung und Bewährungshilfe etwa die Hälfte der Befragten keiner geregelten Beschäftigung nachgeht. Nur 15 % haben einen Arbeitsplatz gefunden und ein Drittel befand sich in einer Ausbildungsmaßnahme. Alkohol und Drogen sind für etwa ein Drittel der Befragten ein schwerwiegendes Problem, da sie sich entweder als abhängig bezeichnen oder täglich bzw. häufig konsumieren. Als Fazit muss also eine durch geringe Arbeitsmarktintegration und hohe Suchtproblematik gekennzeichnete Lebenssituation junger Haftentlassener festgestellt werden, die sich auf eine Legalbewährung vermutlich eher negativ auswirkt. Der Aspekt des Zusammenhangs von Lebenssituation und Legalbewährung wird in der folgenden Untersuchung näher beleuchtet.

3.3.2 Soziale Integration und Delinquenz nach Haftentlassung

Anhand von 1135 persönlichen Interviews in Freiheit wurde versucht, ein genaueres Bild der Lebensumstände entlassener Jugendstrafgefangener zu einem im Vergleich zur vorangegangenen Untersuchung späteren Zeitpunkt zu zeichnen (Lauterbach, 2009). Durchschnittlich zwei Jahre nach ihrer Haftentlassung hat sich die Lebenssituation der ehemaligen Inhaftierten eher verschlechtert als verbessert: Die Mehrheit der Befragten (56 %) geht keiner geregelten Arbeit nach, ein Drittel ist hoch verschuldet und etwa ein Achtel berichtet einen problematischen Alkoholkonsum, zudem konsumiert jeder vierte Befragte häufig illegale Drogen. Ein großer Teil der Befragten ist zudem weiterhin delinquent. Mehr als 40 % geben an, innerhalb der letzten drei Monate ein beliebiges Delikt (ohne Berücksichtigung von Bagatelldelikten wie Kaufhausdiebstahl oder Leistungerschleichung) begangen zu haben und auch in der offiziellen Statistik weisen mehr als 40 % innerhalb von 12 Monaten nach dem Interview Einträge auf. Noch verschärft wird dieser Befund durch die Tatsache, dass die Nachentlassungstichprobe mit hoher Wahrscheinlichkeit Selektionsprozessen unterliegt. Befragte, die bereits kurz nach ihrer Entlassung aufgrund eines Rückfalls reinhaftiert wurden (etwa 38 % innerhalb von 12 Monaten) standen nicht für ein Interview in Freiheit zur Verfügung. Zusätzlich kann vermutet werden, dass Teilnehmer mit häufig wechselndem oder ohne festen Wohnsitz kaum zu erreichen waren und dass diese beiden Gruppen womöglich noch schwierigere Lebensumstände aufweisen als die beschriebenen.

Die in dieser Untersuchung vorrangig verfolgte Fragestellung liegt im Zusammenhang von Merkmalen der sozialen Integration (z.B. Arbeit, Familie, Peerkontakte) und erneuter Delinquenz. Hier zeigte sich sowohl in der bivariaten als auch in der multivariaten Analyse, dass fehlende Erwerbstätigkeit, Substanzkonsum und Kontakte zu Gleichaltrigen mit erhöhter Rückfälligkeit einhergehen. Ohne Einfluss auf die Fortsetzung delinquenten Verhaltens waren dagegen Verschuldung, Familiengründung und Wohnsituation. Die Befunde sind insofern nicht trivial, als dass beispielsweise die Beziehung von Arbeitslosigkeit und Delinquenz weniger eindeutig ist als gemeinhin angenommen. Es finden sich auf der Makroebene keine Zusammenhänge (Eisenberg, 2004; Schumann, 2003) und auch in der Bremer Schulabgänger-Längsschnittstudie fanden sich nur unter bestimmten Konstellationen Einflüsse der Erwerbssituation auf das Legalverhalten (Grobbin, Prein, Reyels & Seus, 2003). Schumann (2004) nimmt deshalb an, dass die Verbindung von Erwerbsstatus mit offiziell registrierter Delinquenz auf der Gleichsetzung von Arbeitsmoral und allgemeiner Moral durch die strafrechtlichen Instanzen und somit auf Sanktionierungstendenzen beruht. Vor diesem Hintergrund kann festgestellt werden, dass der Grad der beruflichen Einbindung ehemaliger Straftäter nachweis-

lich Einfluss auf ihr weiteres Legalverhalten hat. Ein selektiver Einfluss der Strafverfolgung kann in diesem Fall ausgeschlossen werden, da der Effekt sowohl bei offiziell registrierter als auch bei selbstberichteter Delinquenz zu beobachten ist. Zusammenfassend betrachtet sind somit Erwerbstätigkeit, Substanzkonsum und Art der Peerkontakte Merkmale, die mit dem Ausstieg aus kriminellen Karrieren in Zusammenhang stehen und sich aufgrund ihrer Relevanz für das Legalverhalten als Erfolgskriterien strafvollzuglicher Behandlung anbieten.

4 Fazit und Ausblick

Aus den vorliegenden Untersuchungen bleiben drei Hauptergebnisse festzuhalten. Erstens erweisen sich etablierte psychodiagnostische Instrumente im Einsatz mit Straftätern unter bestimmten Voraussetzungen als problematisch und genügen nicht den üblichen Gütekriterien psychologischer Diagnostik. Zweitens ergibt sich eine Diskrepanz zwischen der von Vertretern evidenzbasierter Kriminalpolitik geforderten Güte von Untersuchungsdesigns und den konkreten Bedingungen im Strafvollzug, die Kompromisse hinsichtlich der Designgüte und der praktischen Gegebenheiten erfordern. Drittens schränkt sowohl die alleinige Verwendung offiziell registrierter Rückfälligkeit als auch die alleinige Berücksichtigung kriminogener Risikofaktoren als Effektkriterien von Behandlung die Aussagekraft der Ergebnisse ein. Im Folgenden werden diese drei Ergebnisse im Hinblick auf ihre Konsequenzen für den Jugendstrafvollzug und die Behandlungsforschung in Strafvollzug diskutiert.

4.1 Perspektiven für den Jugendstrafvollzug

Für den Jugendstrafvollzug sind aus den Ergebnissen praktische Konsequenzen für die Anwendung standardisierter Persönlichkeitsdiagnostik in der Behandlungsuntersuchung abzuleiten. Im Einsatz mit Straftätern erwiesen sich etablierte Instrumente zur Erfassung von Persönlichkeitsmerkmalen als anfällig für Antworttendenzen im Sinne sozialer Erwünschtheit und Verfälschung. Da allerdings nicht jedem Befragten ein solches Antwortverhalten im selben Ausmaß zu unterstellen ist und der Grad der Abweichung deshalb individuell variieren kann, stellt die Verwendung eines konstanten Korrekturfaktors keine Lösung dar. Weil auch durch Lügen- oder Offenheitsskalen diese Tendenzen nicht vollständig kontrolliert werden können ist folglich die Zweckmäßigkeit der Erhebung von persönlichkeitsbezogenen Selbstauskünften in einem Zwangskontext ernsthaft in Frage zu stellen. Wenn in der strafvollzuglichen Praxis dennoch standardisierte Persönlichkeitsdiagnostik durchgeführt wird, sollten die Testergebnisse nur verwendet werden, wenn sie sich multimethodal, also beispielsweise durch das Hinzuziehen biografischer Daten oder durch Verhaltensbeobachtungen bestätigen lassen. Auch ist zu überprüfen, inwieweit Rückfallprognoseinstrumente wie etwa die PCL-R oder das

HCR-20, die unter Zuhilfenahme kriminal- und lebensgeschichtlicher Merkmale der Straftäter eine Rückfallvorhersage erlauben (vgl. Dahle, 2005, 2007; Nedopil, 2006; Webster & Bailes, 2001), verstärkt im Rahmen der Behandlungsuntersuchung eingesetzt werden sollten und eine fehlerbehaftete Persönlichkeitsdiagnostik möglicherweise korrigieren können. Zwar werden in der neuen Generation dieser Instrumente als dynamische Risikofaktoren ebenfalls Persönlichkeitseigenschaften mitberücksichtigt, diese werden aber in erster Linie im Rahmen klinischer Interviews und persönlicher Exploration und nicht mithilfe standardisierter Breitband-Diagnostik erhoben. Inwieweit Selbstbeurteilungsverfahren für die Erhebung spezifischer im Rahmen der Prognoseinstrumente relevanter Persönlichkeitsaspekte zusätzlich eingesetzt werden, sollte im Einzelfall entschieden werden.

Weiterhin sollten Verantwortliche im Strafvollzug für die Erfordernisse einer aussagekräftigen und verlässlichen Maßnahmenevaluation sensibilisiert werden, wozu insbesondere die Rekrutierung von Vergleichsgruppen gehört. Da die Verwirklichung echter experimenteller Designs in der Behandlungsevaluation vermutlich weiterhin die Ausnahme bleiben wird, wäre außerdem eine verbesserte Dokumentation von Behandlungsteilnahme und -abbrüchen in den Gefangenenpersonalakten ein leicht umsetzbarer Schritt in diese Richtung. Allein durch die Betrachtung von Behandlungsteilnahme und Rückfälligkeit wird man jedoch allenfalls formal dem gesetzlichen Evaluationsauftrag gerecht, der nur maßnahmenspezifisch unter Einbezug unterschiedlicher Effektivitätskriterien von Behandlung sowie der Kontrolle individueller Teilnehmervoraussetzungen als Prozessmerkmale angemessen zu erfüllen ist. Für eine fundierte und unabhängige Maßnahmenevaluation ist darüber hinaus die Kooperation mit externen Forschungseinrichtungen anzuraten, unter anderem um bei Inhaftiertenbefragungen eine Erhebung durch unabhängige Interviewer zu garantieren und Antworttendenzen soweit möglich zu vermeiden.

Deutlich wird auch, dass prospektiv angelegte Evaluationsstudien einzelner Behandlungsmaßnahmen sich mitunter über mehrere Jahre erstrecken, um eine angemessene Stichprobengröße zu erreichen. So betrug die Behandlungskapazität des LoGo-Trainings 16 Teilnehmer pro Jahr und die Anzahl der in der sozialtherapeutischen Abteilung behandelten Insassen lag vor allem aufgrund der relativen Seltenheit von Sexualdelikten bei Jugendlichen und Heranwachsenden noch deutlich darunter. Eine Schwierigkeit besteht in diesem Zusammenhang in der Maßnahmenintegrität, da die meisten Behandlungen nicht einem unveränderlichen Curriculum folgen, sondern in ihren Behandlungselementen und -methoden fortwährend weiterentwickelt werden, entweder aus inhaltlichen Überlegungen oder weil die Behandlungsteams wechseln. Bei Einbezug des Legalbewährungskriteriums anhand offiziell regist-

rierter Rückfalldaten ist zudem ein angemessener Beobachtungszeitraum von mindestens einem Jahr, besser von mehreren Jahren zu realisieren. Aus diesem Grunde sollte Verantwortlichen im Strafvollzug bewusst gemacht werden, dass Behandlungsevaluation ein zeitaufwändiges Unterfangen darstellt, das ad hoc und unter wechselnden Bedingungen keine verlässlichen Ergebnisse erbringen kann.

4.2 Perspektiven für die Behandlungsforschung im Strafvollzug

Im Forschungskontext können bei der standardisierten Erhebung von Selbstangaben zur Persönlichkeit Antworttendenzen zwar nicht völlig ausgeschlossen werden, sie sind jedoch nicht im selben Ausmaß zu befürchten wie in der Behandlungsuntersuchung. Dies ist in erster Linie darauf zurückzuführen, dass Erhebungen hier anonym oder zumindest glaubwürdig vertraulich durchgeführt werden können und für die Inhaftierten in den meisten Fällen keine direkten Konsequenzen aus der Testung zu erwarten sind. Weiterhin werden Aussagen in erster Linie auf Gruppenebene getroffen, so dass die individuelle Diagnose oder Klassifikation unter Bezug auf Normwerte nicht vorrangiges Untersuchungsziel ist. Hinsichtlich der Validität der Instrumente kann insbesondere auf die differentielle Validität hinsichtlich der Unterscheidung verschiedener Deliktgruppen und der prognostischen Validität bezüglich offiziell registrierter Rückfälligkeit verwiesen werden. Allerdings sollte je nach Kontext der Untersuchung kritisch hinterfragt werden, ob beispielsweise im Rahmen einer Maßnahmenevaluation Validitätsbedrohungen durch Reaktivität der Teilnehmer möglich sind. So ist es denkbar, dass Teilnehmer einer Maßnahme bei offensichtlich behandlungsrelevanten Persönlichkeitsmerkmalen (wie z.B. Aggressivität bei einem Anti-Gewalt-Training) aus Loyalität oder Unzufriedenheit mit den Trainingsdurchführenden mehr oder weniger sozial erwünscht oder verfälschend antworten. Auch wenn dieser Befürchtung gelegentlich mit dem Argument begegnet wird, die Teilnehmer hätten keinerlei „emotionalen Bezug zu derartigen Fragestellungen wie Wissenschaft und Forschung“ (Schanzenbäcker, 2003, S. 140), so wird ein emotionaler Bezug zu den Trainingsdurchführenden oft als Voraussetzung erfolgreicher Behandlung angesehen und könnte demzufolge bei der Erhebung von Selbstauskünften einen Störfaktor darstellen. Unabhängig von möglichen Antworttendenzen können allerdings Verständnisprobleme aufgrund hoher sprachlicher und kognitiver Anforderungen sowie für die Zielgruppe alltagsferne Iteminhalte zu Einschränkungen der Validität der erhobenen Konstrukte führen. Diesem kann nur durch eine systematische Inspektion und Überarbeitung von Iteminhalten und -formulierungen und einer Validierung der überarbeiteten Instrumente für die Zielpopulation begegnet werden. Bei der Verwendung überprüfter und gegebenenfalls angepasster Instru-

mente sind dann allerdings keine grundsätzlichen Bedenken gegen die Verwendung standardisierter Persönlichkeitsdiagnostik im Forschungskontext anzumelden.

Um gesicherte Erkenntnisse über die Wirksamkeit strafvollzuglicher Interventionen und Behandlungsmaßnahmen zu erlangen, sind vorrangig solche Untersuchungsdesigns zu realisieren, die Gefährdungen der internen Validität weitestgehend vermeiden. Idealerweise sollten also echte experimentelle Designs zur Wirksamkeitsüberprüfung von Behandlungsmaßnahmen verwendet werden. Diesem Vorgehen werden allerdings vollzugspraktische, rechtliche und ethische Aspekte entgegengehalten. So ist es für Evaluationsforscher oft schwierig, die Verantwortlichen bei Vorliegen einer entsprechenden Indikation von der Notwendigkeit einer randomisierten Behandlungszuweisung zu überzeugen, da mitunter auch gruppenspezifische Aspekte bei der Zusammenstellung von Behandlungsgruppen eine Rolle spielen. Für bestimmte Deliktgruppen wie Sexualstraftäter scheidet ein solches Verfahren aus rechtlichen Gründen aus, zudem wäre es ethisch fragwürdig, einem rückfallgefährdeten Gewalt- oder Sexualstraftäter aus wissenschaftlichen Gründen die Behandlung zu verweigern. Lässt sich also wie in den meisten Fällen ein randomisiertes Design nicht realisieren, so besteht die nächst niedrigere Stufe der Designqualität nach Sherman et al. (1997) in der Durchführung quasiexperimenteller Untersuchungen mit Kontrolle möglicher Störfaktoren. Es könnte beispielsweise versucht werden, eine Gruppe behandelter Straftäter einer Gruppe unbehandelter Straftäter gegenüberzustellen und dabei bekannte kriminogene Merkmale im Rahmen multivariater Analyseverfahren zu kontrollieren. Weiterhin könnten durch ein Propensity score matching (Becker & Ichino, 2002; Rosenbaum & Rubin, 1983) aus einer Gruppe Unbehandelter statistische Zwillinge ausgewählt werden, die sich in bekannten Risikofaktoren von Delinquenz möglichst gering von den Behandelten unterscheiden. Einschränkungen dieser Verfahren liegen aber in der notwendigen Kenntnis und Erfassung möglicher Störfaktoren sowie der Verfügbarkeit potentieller Kontrollprobanden. Sind Kontrollprobanden nicht verfügbar, so sollte deshalb nicht auf eine Maßnahmenevaluation verzichtet werden, auch wenn der Fokus sich damit vom Wirkungsnachweis zur begleitenden oder maßnahmenoptimierenden Evaluation verschiebt (summative versus formative Evaluation, Scriven, 1972). Neben der Verwendung von Prä-Post-Designs ohne Kontrollgruppe besteht eine mögliche Evaluationsstrategie in der Identifikation von Subgruppen, die besonders gut oder besonders wenig auf eine Behandlung ansprechen (vgl. Peck, 2005).

Weiterhin sollten strafvollzugliche Interventionsmaßnahmen weder allein am Rückfallkriterium, noch allein an der Erreichung von Nahzielen bewertet werden, da beide Kriterien für sich allein genommen nur eingeschränkte Aussagekraft besitzen. Das Rückfallkriterium ist für

die Behandlungsevaluation in vielen Fällen zu insensitiv, unter anderem da neben der erfolgten Behandlung eine Reihe unkontrollierbarer Umweltfaktoren Einfluss auf das beobachtbare Legalverhalten entlassener Straftäter nehmen. Weiterhin erlaubt seine alleinige Verwendung nicht die im Rahmen einer theoriegeleiteten Evaluation zu fordernde Überprüfung einzelner Programmelemente, was nur durch die Erfassung von Prozessmerkmalen und spezifischen Nah- und Zwischenzielen von Behandlung möglich ist. Allein die Untersuchung programmspezifischer Nahziele hingegen lässt den empirischen Nachweis offen, dass eine gegebenenfalls erfolgte Veränderung kriminogener Risikofaktoren tatsächlich die Legalbewährung beeinflusst. Dies ist insbesondere von Bedeutung, wenn als Nahziele psychologische Merkmale erhoben werden, deren Konstruktvalidität nicht immer uneingeschränkt angenommen werden kann. Erst die Erhebung und gleichzeitige Berücksichtigung verschiedener Effektivitätskriterien erlaubt eine umfassende Bewertung der untersuchten Maßnahme. Dabei sollten in Abhängigkeit von den Behandlungszielen auch Aspekte der Sozialbewährung bzw. sozialen Integration in Betracht gezogen werden, die als Effektkriterien häufig eine bessere inhaltliche Verknüpfung zur Behandlungsmaßnahme besitzen und in der vermuteten Kausalkette von Behandlungswirksamkeit ein proximaleres Maß darstellen als die Legalbewährung. Als solche Kriterien bieten sich beispielsweise die Integration in den Arbeitsmarkt, der Substanzkonsum und die Zusammensetzung der Peergruppe an. Ein Grund, warum dies in der Praxis noch relativ selten geschieht liegt in der Betreuungslücke zwischen Haftentlassung und nachsorgenden Diensten und Einrichtungen und der damit verbundenen schwierigen Erreichbarkeit der Entlassenen für die Erhebung entsprechender Daten. Im Zuge einer ohnehin notwendigen Vernetzung dieser Dienste im Sinne eines Übergangsmanagements sollten sich aber Wege finden lassen, entsprechende Daten für die Behandlungsevaluation zur Verfügung zu stellen.

Literatur

- Andrews, D. A., Bonta, J. L. & Hoge, R. D. (1990). Classification for effective rehabilitation: Rediscovering psychology. *Criminal Justice & Behavior*, 17, 19-52.
- Andrews, D. A., Zinger, I., Hoge, R. D., Bonta, J., Gendreau, P. & Cullen, F. (1990). Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 28, 369-404.
- Barbaree, H. (1997). Evaluating treatment efficacy with sexual offenders: The insensitivity of recidivism studies to treatment. *Sexual Abuse: A Journal of Research and Treatment*, 9, 111-128.
- Basqué, M. (1994). Praxis der Delinquenzbehandlung. In M. Steller, K.-P. Dahle & M. Basqué (Hrsg.), *Straftäterbehandlung* (S. 103-112). Pfaffenweiler: Centaurus.
- Becker, S. O. & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal*, 2, 358-377.
- Behnke, M. (2004). Behandlung und Behandlungsplanung. In W. Pecher (Hrsg.), *Justizvollzugspsychologie in Schlüsselbegriffen* (S. 26-39). Stuttgart: Kohlhammer.
- Behnke, M. & Endres, J. (2008). Behandlung im Strafvollzug. In M. Steller & R. Volbert (Hrsg.), *Handbuch der Rechtspsychologie* (S. 107-118). Göttingen: Hogrefe.
- Bereswill, M., Koesling, A. & Neuber, A. (2007). Brüchige Erfolge - Biographische Diskontinuität, Inhaftierung und Integration. In J. Goerdeler & P. Walkenhorst (Hrsg.), *Jugendstrafvollzug in Deutschland: Neue Gesetze, neue Strukturen, neue Praxis? Schriftenreihe der DVJJ Bd. 40* (S. 294-312). Mönchengladbach: Forum Verlag Godesberg.
- Bertram, C. (2004). Wider den organisierten Beziehungsabbruch. Entlassungsvorbereitung als kontinuierliches Hilfsangebot in einem vernetzten System. In G. Rehn, R. Nanninga & A. Thiel (Hrsg.), *Freiheit und Unfreiheit. Arbeit mit Straftätern innerhalb und außerhalb des Justizvollzugs* (S. 430-446). Herbolzheim: Centaurus.
- Biedermann, T. (2007). *Beiträge zur Aggressionsdiagnostik: Multimethodale psychologische Informationsgewinnung bei jungen inhaftierten Gewalttätern im Vergleich mit sozial unauffälligen Jugendlichen*. Hamburg: Kovač.
- Bliesener, T. (2008). Prävention und Bewältigung von Delinquenz und Devianz. In F. Petermann & W. Schneider (Hrsg.), *Enzyklopädie der Psychologie: Angewandte Entwicklungspsychologie* (S. 677-719). Göttingen: Hogrefe.
- Bliesener, T. & Köhnken, G. (2005). Kriminalpsychologie. In D. Frey & C. Graf Hoyos (Hrsg.), *Psychologie in Gesellschaft, Kultur und Umwelt* (S. 11-17). Weinheim: Beltz.

- Borkenau, P. & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study of the five-factor model of personality. *Personality and Individual Differences, 11*, 515-524.
- Boruch, R. F. (1997). *Randomized experiments for planning and evaluation: A practical guide*. Thousand Oaks: Sage.
- Bosold, C. (2008). Therapieindikation bei der Straftäterbehandlung. In M. Steller & R. Volbert (Hrsg.), *Handbuch der Rechtspsychologie* (S. 144-151). Göttingen: Hogrefe.
- Bosold, C., Hosser, D. & Lauterbach, O. (2007). Psychosoziale Behandlung im Jugendstrafvollzug: Erste Ergebnisse einer Längsschnittstudie. *Praxis der Rechtspsychologie, 17*, 265-284.
- Bosold, C. & Lauterbach, O. (2010). Leben ohne Gewalt organisieren: Evaluation eines Trainings für Gewalttäter im Jugendstrafvollzug. *Forensische Psychiatrie, Psychologie, Kriminologie, 4*, 267-277.
- Bosold, C., Lauterbach, O. & Hosser, D. (2006). *Evaluation eines kognitiv-behavioralen Gruppentrainings für Gewalttäter*. Vortrag auf dem 45. Kongress der Deutschen Gesellschaft für Psychologie, 17.-21. September in Nürnberg.
- Bosold, C., Prasse, A. & Lauterbach, O. (2006). Anti-Gewalttrainings im Jugendvollzug. Eine bundesweite Bestandsaufnahme. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe, 17*, 27-37.
- Boxberg, V. & Bosold, C. (2009). Soziales Training im Jugendstrafvollzug: Effekte auf die Sozial- und Legalbewährung. *Forensische Psychiatrie, Psychologie, Kriminologie, 3*, 237-243.
- Brody, S. R. (1976). *The effectiveness of sentencing: A review of the literature* (Home Office Research Study No. 35). London: Home Office.
- Campbell, D. T. (1969). Reforms as experiments. *American Psychologist, 24*, 409-429
- Chen, H.-T. & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. *Evaluation Review, 7*, 283-302.
- Church, A. T. & Burke, P. J. (1994). Exploratory and confirmatory tests of the big five and Tellegen's three- and four-dimensional models. *Journal of Personality and Social Psychology, 66*, 93-114.
- Coester, M., Bannenberg, B. & Rössner, D. (2007). Die deutsche kriminologische Evaluationsforschung im internationalen Vergleich. In F. Lösel, D. Bender & J.-M. Jehle (Hrsg.), *Kriminologie und wissenschaftsbasierte Kriminalpolitik: Entwicklungs- und Evaluationsforschung* (S. 93-112). Mönchengladbach: Forum Verlag Godesberg.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design issues for field settings*. Chicago: McNally.
- Cornel, H. (1995). Resozialisierung - Klärung des Begriffs, seines Inhalts und seiner Verwendung. In H. Cornel, B. Maelicke & B. R. Sonnen (Hrsg.), *Handbuch Resozialisierung* (S. 13-53). Baden-Baden: Nomos.
- Cottonaro, S. & Suhling, S. (2007). Weichenstellung im Betreuungskontinuum: Diagnose, Prognose, Indikation und Vollzugsplanung. In J. Goerdeler & P. Walkenhorst (Hrsg.), *Jugendstrafvollzug in Deutschland: Neue Gesetze, neue Strukturen, neue Praxis?* (S. 222-237). Mönchengladbach: Forum Verlag Godesberg.
- Cullen, F. T. & Gendreau, P. (2001). From nothing works to what works: Changing professional ideology in the 21st century. *The Prison Journal*, 81, 313-338.
- Dahle, K.-P. (1997). Therapie und Therapieindikation bei Straftätern. In M. Steller & R. Volbert (Hrsg.), *Psychologie im Strafverfahren* (S. 142-159). Bern: Huber.
- Dahle, K.-P. (2005). Psychologische Begutachtung zur Kriminalprognose. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren: Indikationen, Methoden und Qualitätsstandards* (S. 133-169). Darmstadt: Steinkopff.
- Dahle, K.-P. (2007). Methodische Grundlagen der Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 101-110.
- Dahle, K.-P., Schneider, V. & Ziethen, F. (2007). Standardisierte Instrumente zur Kriminalprognose. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 15-26.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 85.
- Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113-126.
- Davis, M. H. & Franzoi, S. L. (1991). Stability and change in adolescent self-consciousness and empathy. *Journal of Research in Personality*, 25, 70-87.
- Dowden, C. & Andrews, C. A. (1999). What works in young offender treatment: A meta-analysis. *Forum on Corrections Research*, 11(2), 21-24.
- Dünkel, F. (1980). *Legalbewährung nach sozialtherapeutischer Behandlung. Eine empirische und vergleichende Untersuchung anhand der Strafregisterauszüge von 1503 in den Jahren 1971-1974 entlassenen Strafgefangenen in Berlin Tegel*. Berlin: Duncker und Humblot.

- Dünkel, F. (2005). Reformen des Sexualstrafrechts und Entwicklung der Sexualdelinquenz in Deutschland. In D. Schläfke, F. Hässler & J. M. Fegert (Hrsg.), *Sexualstraftaten: Forensische Begutachtung, Diagnostik und Therapie* (S. 1-31). Stuttgart: Schattauer.
- Dünkel, F. & Geng, B. (2007). Rechtstatsächliche Befunde zum Jugendstrafvollzug in Deutschland. *Forum Strafvollzug - Zeitschrift für Strafvollzug und Straffälligenhilfe*, 56, 65-80.
- Egg, R. (Hrsg.). (2000). *Behandlung von Sexualstraftätern im Justizvollzug: Folgerungen aus den Gesetzesänderungen*. Wiesbaden: Kriminologische Zentralstelle.
- Egg, R. (2004). Rückfalluntersuchungen anhand von Bundeszentralregisterauskünften - am Beispiel von Sexualstraftätern. In W. Heinz & J.-M. Jehle (Hrsg.), *Rückfallforschung* (S. 119-130). Wiesbaden: Kriminologische Zentralstelle.
- Egg, R. (2007). Rückfälligkeit nach Straf- und Maßregelvollzug. In F. Lösel, D. Bender & J.-M. Jehle (Hrsg.), *Kriminologie und wissenschaftsbasierte Kriminalpolitik: Entwicklungs- und Evaluationsforschung* (S. 247-267). Mönchengladbach: Forum Verlag Godesberg.
- Eisenberg, N. & Strayer, J. (1987). Critical issues in the study of empathy. In N. Eisenberg & J. Strayer (Eds.), *Empathy and its development* (pp. 3-13). Cambridge: Cambridge University Press.
- Eisenberg, U. (2004). Arbeitslosigkeit und Delinquenz. In U. Baßeler, M. Heintzen & L. Kruschwitz (Hrsg.), *Arbeitslosigkeit: Ringvorlesung der Fachbereiche Rechts- und Wirtschaftswissenschaft der Freien Universität Berlin* (S. 67-86). Berlin: Duncker und Humblot.
- Endrass, J., Rossegger, A., Noll, T. & Urbaniok, F. (2008). Wirksamkeit von Therapien bei Gewalt- und Sexualstraftätern. *Psychiatrische Praxis*, 35, 8-14.
- Enzmann, D. (1996). *Gestresst, erschöpft oder ausgebrannt? Einflüsse von Arbeitssituation, Empathie und Coping auf den Burnoutprozess*. München: Profil.
- Eysenck, H. J. (1965). *Crime and personality*. Norwich: Fletcher & Son.
- Fahrenberg, J., Hampel, R. & Selg, H. (2001). *FPI-R: Das Freiburger Persönlichkeitsinventar (7., überarbeitete und neu normierte Auflage)*. Göttingen: Hogrefe.
- Farrington, D. P., Gottfredson, D. C., Sherman L. W. & Welsh, B. C. (2002). The Maryland scientific methods scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (pp. 13-21). London: Routledge.
- Feltes, T. (1991). Rückfall. In H.-J. Kerner (Hrsg.), *Kriminologie Lexikon* (S. 279-281). Heidelberg: Kriminalistik Verlag.

- Gendreau, P., Little, T. & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology*, 34, 575-607.
- Gollwitzer, M. (2007). Ansätze zur Primär- und Sekundärprävention aggressiven Verhaltens bei Kindern und Jugendlichen. In Gollwitzer, M., Pfetsch, J., Schneider, V., Schultz, A., Steffke, T. & Ulrich, C. (Hrsg.), *Gewaltprävention bei Kindern und Jugendlichen: Aktuelle Erkenntnisse aus Forschung und Praxis* (S. 141-157). Göttingen: Hogrefe.
- Graebisch, C. (2004). Evidence-based crime prevention. *Kriminologisches Journal*, 36, 266-283.
- Greve, W. (1998). *The consequences of prisonization for juveniles: A theoretical and methodological framework for research* (KFN Forschungsberichte Nr. 74). Hannover: KFN.
- Greve, W. (2004). Die Entwicklungsfolgen der Jugendstrafe. In H. Schöch & J.-M. Jehle (Hrsg.), *Angewandte Kriminologie zwischen Freiheit und Sicherheit* (S. 157-171). Mönchengladbach: Forum Verlag Godesberg.
- Greve, W. & Hosser, D. (1998). Psychische und soziale Folgen einer Jugendstrafe: Forschungsstand und Desiderate. *Monatsschrift für Kriminologie und Strafrechtsreform*, 81, 83-103.
- Greve, W., Hosser, D. & Pfeiffer, C. (1997). *Gefängnis und die Folgen. Identitätsentwicklung und kriminelles Handeln während und nach Verbüßung einer Jugendstrafe. Konzeption und Planung eines Forschungsprojektes* (KFN Forschungsberichte Nr. 64). Hannover: KFN.
- Grietens, H. & Hellinckx, W. (2004). Evaluation effects of residential treatment for juvenile offenders by statistical metaanalysis: A review. *Aggression and Violent Behavior*, 9, 401-415.
- Grobbin, A., Prein, G., Reyels, W. & Seus, L. (2003). Nach der zweiten Schwelle: Zur Bedeutung der beruflichen Etablierung im Lebensverlauf. In K. F. Schumann (Hrsg.), *Berufsbildung, Arbeit und Delinquenz: Bremer Längsschnittstudie zum Übergang von der Schule in den Beruf bei ehemaligen Hauptschülern* (Bd. 1, S. 175-212). Weinheim: Juventa.
- Gudjonsson, G. H. (2001). Psychometric assessment. In C. R. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 111-122). Chichester: Wiley.
- Hager, W., Leichsenring, F. & Schiffler, A. (1999). Evaluationsparadigmen: Zur Bedeutung der Unterscheidung von vergleichenden und isolierten Evaluationen in der Psychotherapieforschung. *Psychotherapeut*, 44, 234-240.

- Hanson, R. K. (1997). How to know what works with sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 9, 129-145.
- Hare, R. D. (1980). A research scale for the assessment of psychopathy in criminal populations. *Personality and Individual Differences*, 1, 111-119.
- Harkins, L. & Beech, A. (2007). Measurement of the effectiveness of sex offender treatment. *Aggression and Violent Behavior*, 12, 36-44.
- Harrendorf, S. (2006). Gefährliche Gewalttäter? Ergebnisse einer bundesweiten Rückfallstudie. *Bewährungshilfe*, 308-338.
- Havighurst, R. J. (1972). *Developmental tasks and education*. New York: McKay.
- Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior*, 16, 257-272.
- Heilemann, M., Fischwasser-von Proeck, G. (2001). *Gewalt wandeln: Das Anti-Aggressivitäts-Training AAT*. Lengerich: Pabst.
- Heinz, W. (1998). Kriminalprävention: Anmerkungen zu einer überfälligen Kurskorrektur der Kriminalpolitik. In Kerner, H.-J., Jehle, J.-M. & Marks, E. (Hrsg.), *Entwicklung der Kriminalprävention in Deutschland: Allgemeine Trends und bereichsspezifische Perspektiven* (S. 17-59). Mönchengladbach: Forum Verlag Godesberg.
- Heinz, W. (2006). Zum Stand der Dunkelfeldforschung in Deutschland. In J. Obergfell-Fuchs & M. Brandenstein (Hrsg.), *Nationale und internationale Entwicklungen in der Kriminologie: Festschrift für Helmut Kury zum 65. Geburtstag* (S. 241-263). Frankfurt: Verlag für Polizeiwissenschaft.
- Heinz, W. & Jehle, J.-M. (Hrsg.). (2004). *Rückfallforschung*. Wiesbaden: Kriminologische Zentralstelle.
- Höynck, T., Hagemann, N., Kapteina, B.-M., Klimaschewski, K., Lübke, V., Luu, N. & Riechey, F. (2008). Jugendstrafvollzugsgesetze der Länder: Eine Auswahl wichtiger Regelungsbereiche in synoptischer Darstellung. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe*, 19, 159-166.
- Hollis, S. & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*, 319, 670-674.
- Holthusen, B. & Lüders, C. (2003). Evaluation von Kriminalitätsprävention: Eine thematische Einleitung. In Arbeitsstelle Kinder- und Jugendkriminalitätsprävention (Hrsg.), *Evaluierte Kriminalitätsprävention in der Kinder- und Jugendhilfe. Erfahrungen und Ergebnisse aus fünf Modellprojekten* (S. 9-30). München: Deutsches Jugendinstitut.

- Hope, T. (2005). Pretend it doesn't work: The 'anti-social' bias in the Maryland Scientific Methods Scale. *European Journal on Criminal Policy and Research*, 11, 275-296.
- Hosser, D. (2008). Prisonisierungseffekte. In M. Steller & R. Volbert (Hrsg.), *Handbuch der Rechtspsychologie* (S. 172-179). Göttingen: Hogrefe.
- Hosser, D. & Bosold, C. (2006). *Entwicklungsfolgen der Jugendstrafe: Eine längsschnittliche Untersuchung von erstmals inhaftierten Jugendlichen und Heranwachsenden*. Fortsetzungsantrag an die Deutsche Forschungsgemeinschaft.
- Hosser, D. & Bosold, C. (2008). Behandlung im Jugendvollzug. In M. Steller & R. Volbert (Hrsg.), *Handbuch der Rechtspsychologie* (S. 128-134). Göttingen: Hogrefe.
- Hosser, D., Bosold, C. & Lauterbach, O. (2006). Sozialtherapeutische Behandlung von jungen Sexualstraftätern: Ergebnisse einer Evaluationsstudie. *Recht & Psychiatrie*, 24, 125-133.
- Hosser, D., Lauterbach, O. & Camehn, K. (2008). Validität und Reliabilität des FPI-R beim Einsatz im Strafvollzug. *Diagnostica*, 54, 129-137.
- Hosser, D., Lauterbach, O. & Höyneck, T. (2007). Und was kommt danach? Entlassungsvorbereitung und Nachentlassungssituation junger Straftentlassener. In J. Goerdeler & P. Walkenhorst (Hrsg.), *Jugendstrafvollzug in Deutschland: Neue Gesetze, neue Strukturen, neue Praxis? Schriftenreihe der DVJJ Bd. 40* (S. 396-412). Mönchengladbach: Forum Verlag Godesberg.
- Jehle, J.-M. (2007). Methodische Probleme einer Rückfallforschung aufgrund von Bundeszentralregisterdaten. In F. Lösel, D. Bender & J.-M. Jehle (Hrsg.), *Kriminologie und wissenschaftsbasierte Kriminalpolitik: Entwicklungs- und Evaluationsforschung* (S. 227-245). Mönchengladbach: Forum Verlag Godesberg.
- Jolliffe, D. & Farrington, D. P. (2004). Empathy and offending: A systematic review and meta-analysis. *Aggression and Violent Behavior*, 9, 441-476.
- Kerner, H.-J. (1991). Dunkelfeldforschung. In H.-J. Kerner (Hrsg.), *Kriminologie Lexikon* (S. 83-85). Heidelberg: Kriminalistik Verlag.
- Kerner, H.-J. (2004). Soziale Bindungen und soziale Abweichung: Zur Bedeutung von Beziehungsschwäche und Empathiemängeln für schwere Jugenddelinquenz. In G. Klosinski (Hrsg.), *Empathie und Beziehung: Zu den Voraussetzungen, Gefährdungen und Verbesserungen menschlicher Beziehungsfähigkeit* (S. 41-64). Tübingen: Attempto.

- Klein, L., Petran, W. & Weber, J. (2007). Übergangsmanagement beginnt mit der Kompetenzfeststellung - Ein Diagnose-Qualifizierungs- und Reintegrationsprogramm. Erfahrungsbericht eines Modellversuchs in der JVA Wiesbaden. In J. Goerdeler & P. Walkenhorst (Hrsg.), *Jugendstrafvollzug in Deutschland: Neue Gesetze, neue Strukturen, neue Praxis? Schriftenreihe der DVJJ Bd. 40* (S. 450-467). Mönchengladbach: Forum Verlag Godesberg.
- Krahé, B. (1989). Faking personality profiles on a standard personality inventory. *Personality and Individual Differences, 10*, 437-443.
- Krampen, G. (1991). *Fragebogen zu Kompetenz- und Kontrollüberzeugung*. Göttingen: Hogrefe.
- Kubinger, K. D. (2002). On faking personality inventories. *Psychologische Beiträge, 44*, 10-16.
- Kury, H. (Hrsg.). (1983a). *Methodische Probleme der Behandlungsforschung - insbesondere in der Sozialtherapie*. Köln: Carl Heymanns.
- Kury, H. (1983b). Verfälschungstendenzen bei Persönlichkeitsfragebogen im Strafvollzug. *Monatsschrift für Kriminologie und Strafrechtsreform, 66*, 72-74.
- Kury, H. (1983c). Zur Methodendiskussion in der Behandlungsforschung. In H. Kury (Hrsg.), *Methodische Probleme der Behandlungsforschung - insbesondere in der Sozialtherapie* (S. 27-79). Köln: Carl Heymanns.
- Kury, H. (1986). *Die Behandlung Straffälliger. Teilband 1: Inhaltliche und methodische Probleme der Behandlungsforschung*. Berlin: Duncker und Humblot.
- Kury, H. (2002). Das Freiburger Persönlichkeitsinventar und sein Einsatz bei kriminologischen Fragestellungen. Das Problem der Verfälschungstendenzen. In M. Myrtek (Hrsg.), *Die Person im biologischen und sozialen Kontext* (S. 248-270). Göttingen: Hogrefe.
- Kury, H. & Beckers, C. (1983). Probleme der Psychodiagnostik bei sozial Auffälligen, insbesondere im Bereich des Strafvollzugs. *Monatsschrift für Kriminologie und Strafrechtsreform, 66*, 63-72.
- Kuschel, A., Hahlweg, K., Miller, Y., Bertram, H., Heinrichs, N. & Saßmann, H. (2003). Universelle Prävention kindlicher Verhaltensstörungen - das Triple P-Programm. In U. Lehmkuhl (Hrsg.), *Aggressives Verhalten bei Kindern und Jugendlichen: Ursachen, Prävention, Behandlung* (S. 175-182). Göttingen: Vandenhoeck & Ruprecht.

- Landenberger, N. A. & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451-476.
- Laubenthal, K. (2003). *Strafvollzug*. Berlin: Springer.
- Lauterbach, O. (2009). Soziale Integration und Delinquenz nach Entlassung aus dem Jugendstrafvollzug. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe*, 20, 44-50.
- Lauterbach, O. & Hosser, D. (2007). Assessing empathy in prisoners: A shortened version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology*, 66, 91-101.
- Leeuw, F. L. (2005). Trends and developments in program evaluation in general and criminal justice programs in particular. *European Journal on Criminal Policy and Research*, 11, 233-258.
- Lipsey, M. W. (1992). The effect of treatment on juvenile delinquents: Results from meta-analysis. In F. Lösel, D. Bender & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 131-143). Berlin: de Gruyter.
- Lipsey, M. W. (1995). What do we learn from 400 research studies on the effectiveness of treatment with juvenile delinquents. In J. McGuire (Eds.), *What works: Reducing reoffending* (pp. 63-111). Chichester: Wiley
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *Annals of the American Academy of Political and Social Science*, 587, 69-81.
- Lipsey, M. W. & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent offenders: Risk factors and successful intervention* (pp. 313-345). Thousand Oaks: Sage.
- Lipton, D. S., Martinson, R. & Wilks, W. (1975). *Effectiveness of correctional treatment: A survey of treatment evaluation studies*. New York: Praeger.
- Littmann, E. (1992). Using psychological tests in the forensic assessment of offenders. In F. Lösel, D. Bender & T. Bliesener (Eds.), *Psychology and law: International perspectives* (pp. 111-120). Berlin: de Gruyter.
- Lösel, F. (1975). Persönlichkeitspsychologische Aspekte delinquenten Verhaltens. In A. Abele, S. Mitzlaff & W. Nowack (Hrsg.), *Abweichendes Verhalten: Erklärungen, Scheinerklärungen und praktische Probleme* (S. 155-177). Stuttgart: Frommann-Holzboog.
- Lösel, F. (1983). Empirische Persönlichkeitsforschung und Delinquenzerklärung. In F. Lösel (Hrsg.), *Kriminalpsychologie* (S. 29-40). Weinheim: Beltz.

- Lösel, F. (1994). Meta-analytische Beiträge zur wiederbelebten Diskussion des Behandlungsgedankens. In M. Steller, K.-P. Dahle & M. Basqué (Hrsg.), *Straftäterbehandlung* (S. 13-34). Pfaffenweiler: Centaurus.
- Lösel, F. (2004). Entwicklungsbezogene und technische Kriminalprävention: Konzeptuelle Grundlagen und Ergebnisse. In H. Schöch & J.-M. Jehle (Hrsg.), *Angewandte Kriminologie zwischen Freiheit und Sicherheit* (S. 175-203). Mönchengladbach: Forum Verlag Godesberg.
- Lösel, F. (2009). Förderung einer evidenz-basierten Politik durch systematische Synthesen: Die Campbell Collaboration. *Psychologische Rundschau*, 60, 246-247.
- Lösel, F. & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *Annals of the American Academy of Political and Social Science*, 587, 84-109.
- Lösel, F. & Bender, D. (1997). Straftäterbehandlung: Konzepte, Ergebnisse, Probleme. In M. Steller & R. Volbert (Hrsg.), *Psychologie im Strafverfahren* (S. 171-204). Bern: Huber.
- Lösel, F., Köferl, P. & Weber, F. (1987). *Meta-Evaluation der Sozialtherapie: Qualitative und quantitative Analysen zur Behandlungsforschung in sozialtherapeutischen Anstalten des Justizvollzugs*. Stuttgart: Enke.
- Lösel, F. & Nowack, W. (1987). Evaluationsforschung. In J. Schultz-Gambard (Hrsg.), *Angewandte Sozialpsychologie: Konzepte, Ergebnisse, Perspektiven* (S. 57-87). München: PVU.
- Lösel, F. & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology*, 1, 117-146.
- MacKenzie, D. L. (1997). Criminal justice and crime prevention. In L. W. Sherman, D. Gottfredson, D. MacKenzie, J. Eck, P. Reuter & S. Bushway (Eds.), *Preventing crime: What works, what doesn't, what's promising. A report to the United States Congress* (pp. 1-76). Department of Criminology and Criminal Justice, University of Maryland.
- Marques, J. K. (1999). How to answer the question "Does sex offender treatment work?". *Journal of Interpersonal Violence*, 14, 437-451.
- Marshall, I. H. (2005). International transferability of the practice of evidence-based crime policies. *European Journal on Criminal Policy and Research*, 11, 225-231.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *Public Interest*, 35, 22-54.

- Martinson, R. (1979). New findings, new views: A note of caution regarding sentencing reform. *Hofstra Law Review*, 7, 243-258.
- McCord, J. (1992). The Cambridge-Somerville study: A pioneering longitudinal experimental study of delinquency prevention. In J. McCord & R. E. Tremblay (Eds.), *Preventing antisocial behavior: Interventions from birth through adolescence* (pp. 196-206). New York: Guilford.
- McCord, J. (2003). Cures that harm: Unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science*, 587, 16-30.
- Miller, P. A. & Eisenberg, N. (1988). The relation of empathy to aggressive and externalizing/antisocial behavior. *Psychological Bulletin*, 103, 324-344.
- Mischel, W. (1961). Preference for delayed reinforcement and social responsibility. *Journal of Abnormal and Social Psychology*, 62, 1-7.
- Müller, K. (2002). Behandeln statt strafen, aber wie? Die Erziehungs- und Behandlungsplanung. In M. Bereswill & T. Höynck (Hrsg.), *Jugendstrafvollzug in Deutschland: Grundlagen, Konzepte, Handlungsfelder* (S. 107-142). Mönchengladbach: Forum Verlag Godesberg.
- Müller-Kohlenberg, H. (2006). Zwei neuralgische Punkte der Selbstevaluation: Unparteilichkeit und Professionalität. In W. Böttcher, H. G. Holtappels & M. Brohm (Hrsg.), *Evaluation im Bildungswesen: Eine Einführung in Grundlagen und Praxisbeispiele* (S. 87-95). Weinheim: Juventa.
- Müller-Marsell, S. (2004). Subkultur im Strafvollzug. In W. Pecher (Hrsg.), *Justizvollzugspsychologie in Schlüsselbegriffen* (S. 286-298). Stuttgart: Kohlhammer.
- Nedopil, N. (2006). *Prognosen in der Forensischen Psychiatrie: Ein Handbuch für die Praxis*. Lengerich: Pabst.
- Ohlemacher, T., Sögding, D., Höynck, T., Ethé, N. & Welte, G. (2001). "Nicht besser, aber auch nicht schlechter": Anti-Aggressivitäts-Training und Legalbewährung. *DVJJ-Journal*, 12, 380-386.
- Ortmann, R. (1987). *Resozialisierung im Strafvollzug: Theoretischer Bezugsrahmen und empirische Ergebnisse einer Längsschnittstudie zu den Wirkungen von Strafvollzugsmaßnahmen*. Freiburg: Max-Planck-Institut für ausländisches und internationales Strafrecht.
- Ortmann, R. (2002). *Sozialtherapie im Strafvollzug*. Freiburg: Max-Planck-Institut für ausländisches und internationales Strafrecht.

- Orr, L. L. (1999). *Social experiments: Evaluating public programs with experimental methods*. Thousand Oaks: Sage.
- Otto, M. (1994). Soziales Training. In M. Steller, K.-P. Dahle & M. Basqué (Hrsg.), *Straftäterbehandlung* (S. 113-131). Pfaffenweiler: Centaurus.
- Pawson, R. & Tilley, N. (1994). What works in evaluation research? *British Journal of Criminology*, 34, 291-306.
- Peck, L. R. (2005). Using cluster analysis in program evaluation. *Evaluation Review*, 29, 178-196.
- Petrosino, A. & Soydan, H. (2005). The impact of program developers as evaluators on criminal recidivism: Results from meta-analyses of experimental and quasi-experimental research. *Journal of Experimental Criminology*, 1, 435-450.
- Petrosino, A., Turpin-Petrosino, C. & Buehler, J. (2003). Scared straight and other juvenile awareness programs for preventing juvenile delinquency: A systematic review of the randomized experimental evidence. *Annals of the American Academy of Political and Social Science*, 589, 41-62.
- Rasch, W. & Kühl, K.-P. (1977). Psychologische Kriterien für die Unterbringung in einer sozialtherapeutischen Anstalt. In W. Rasch (Hrsg.), *Forensische Sozialtherapie* (S. 203-359). Heidelberg: Müller.
- Rasch, W. & Kühl, K.-P. (1978). Psychologische Befunde und Rückfälligkeit nach Aufenthalt in der sozialtherapeutischen Modellanstalt Düren. *Bewährungshilfe*, 25, 44-57.
- Rau, T. (2006a). Die Wirksamkeit des Anti-Aggressivitäts-Trainings lässt sich bestätigen. *Sozial Extra*, 30 (7-8), 38-41.
- Rau, T. (2006b). *Katamnestische Untersuchung zur Wirksamkeit des Anti-Aggressivitäts-Trainings bei straffälligen Jugendlichen und jungen Erwachsenen*. Dissertation, Universität Ulm.
- Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences*, 24, 291-293.
- Repp, N., Köhler, D. & Hinrichs, G. (2004). Psychotherapeutische Angebote in den Jugendanstalten. *Zeitschrift für Strafvollzug und Straffälligenhilfe*, 53, 199-210.
- Rosenbaum, P. R. & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-44.
- Rossi, P. H., Berk, R. A. & Lenihan, K. J. (1980). *Money, work, and crime: Experimental evidence*. New York: Academic Press.

- Rossi, P. H., Berk, R. A. & Lenihan, K. J. (1982). Saying it wrong with figures: A comment on Zeisel. *American Journal of Sociology*, 88, 390-393.
- Sack, F. (1993). Dunkelfeld. In G. Kaiser, H.-J. Kerner, F. Sack & H. Schellhoss (Hrsg.), *Kleines Kriminologisches Wörterbuch* (S. 99-107). Heidelberg: Müller.
- Sarre, R. (1999). *Beyond 'what works?'* A 25 year jubilee retrospective of Robert Martinson. Paper presented at the History of Crime, Policing and Punishment Conference, Canberra, 9-10 December.
- Schanzenbäcker, S. (2003). *Anti-Aggressivitäts-Training auf dem Prüfstand: Gewalttäter-Behandlung lohnt sich*. Herbolzheim: Centaurus.
- Scheurer, H. (1993). *Persönlichkeit und Kriminalität*. Regensburg: Roderer.
- Scheurer, H. & Richter, P. (2005). Psychologische Persönlichkeitsdiagnostik: Zur Bedeutung von Persönlichkeitsfragebogen bei der Begutachtung von Schuldfähigkeit. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren* (S. 39-60). Darmstadt: Steinkopff.
- Schmucker, M. & Lösel, F. (2007). Wie erfolgreich ist die Therapie von Sexualstraftätern? Ergebnisse und Probleme der Wirkungsforschung. In F. Lösel, D. Bender & J.-M. Jehle (Hrsg.), *Kriminologie und wissenschaftsbasierte Kriminalpolitik: Entwicklungs- und Evaluationsforschung* (S. 295-314). Mönchengladbach: Forum Verlag Godesberg.
- Schumann, K. F. (2003). Berufsbildung, Arbeit und Delinquenz: Forschungsstand und offene Fragen. In K. F. Schumann (Hrsg.), *Berufsbildung, Arbeit und Delinquenz: Bremer Längsschnittstudie zum Übergang von der Schule in den Beruf bei ehemaligen Hauptschülern* (Bd. 1, S. 9-44). Weinheim: Juventa.
- Schumann, K. F. (2004). Sind Arbeitsbiographie und Straffälligkeit miteinander verknüpft? Aufklärungen durch die Lebenslaufforschung. *Monatsschrift für Kriminologie und Strafrechtsreform*, 87, 222-243.
- Scriven, M. (1972). Die Methodologie der Evaluation. In C. Wulf (Hrsg.), *Evaluation: Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen* (S. 60-91). München: Piper.
- Seitz, W. (1983). Zur Struktur und Erfassung der Persönlichkeit von Inhaftierten - am Beispiel eines inhaftierungsadäquaten Persönlichkeitsfragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 261-281.
- Seiwald, B. B. (2003). Lügenskala. In K. D. Kubinger & R. S. Jäger (Hrsg.), *Schlüsselbegriffe der psychologischen Diagnostik* (S. 271-273). Weinheim: Beltz.

- Shadish, W. R., Cook, T. D. & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shao, J. & Zhong, B. (2003). Last observation carry-forward and last observation analysis. *Statistics in Medicine*, 22, 2429-2441.
- Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J., Reuter, P. & Bushway, S. D. (Eds.). (1997). *Preventing crime: What works, what doesn't, what's promising. A report to the United States Congress*. Department of Criminology and Criminal Justice, University of Maryland.
- Smith, H. L. (1990). Specification problems in experimental and nonexperimental social research. *Sociological Methodology*, 20, 59-91.
- Statistisches Bundesamt (2009). *Strafvollzug - Demographische und kriminologische Merkmale der Strafgefangenen zum Stichtag 31.3.2008*. Wiesbaden: Statistisches Bundesamt.
- Steller, M. (1994). Behandlung und Behandlungsforschung. In M. Steller, K.-P. Dahle & M. Basqué (Hrsg.), *Straftäterbehandlung* (S. 3-12). Pfaffenweiler: Centaurus.
- Stenzel, M. (1990). *Berufserziehung straffälliger Jugendlicher und Heranwachsender*. Frankfurt am Main: Lang.
- Stieglitz, R.-D. (1993). Selbstbeurteilungsverfahren. In R.-D. Stieglitz & U. Baumann (Hrsg.), *Psychodiagnostik psychischer Störungen* (S. 67-78). Stuttgart: Enke.
- Stock, S. (1993). *Behandlungsuntersuchung und Vollzugsplan: Zum Instrumentarium einer an der Rückfallverhinderung orientierten Ausgestaltung des Strafvollzugs in der Bundesrepublik Deutschland*. Egelsbach: Hänsel-Hohenhausen.
- Streiner, D. L. (2002). The case of the missing data: Methods of dealing with dropouts and other research vagaries. *Canadian Journal of Psychiatry*, 47, 70-77.
- Suhling, S. (2006). Zur Untersuchung der allgemeinen und differentiellen Wirksamkeit sozialtherapeutischer Behandlung im Justizvollzug - Konzepte aus Niedersachsen. *Bewährungshilfe*, 53, 240-259.
- Suhling, S. & Cottonaro, S. (2005). Motivation ist alles? Formen und Bedingungen von Veränderungs- und Behandlungsbereitschaft bei Inhaftierten. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe*, 16, 385-396.
- Trochim, W. M. K. (2001). *The research methods knowledge base*. Cincinnati, OH: Atomic Dog.

- Ullrich, S. (1999). *Die Persönlichkeit von Straftätern: Psychopathologische und normalpsychologische Akzentuierungen*. Dissertation, Martin-Luther-Universität Halle-Wittenberg.
- Walter, J. (2002). Jugendvollzug in der Krise? *DVJJ-Journal*, 13, 127-143.
- Walter, J. (2003). Junge russischsprachige Aussiedler als Klientel in den Justizvollzugsanstalten. In DBH-Bildungswerk (Hrsg.), *Spätaussiedler: Interkulturelle Kompetenz für die Straffälligenhilfe und den Justizvollzugsdienst* (S. 87-120). Mönchengladbach: Forum Verlag Godesberg.
- Webster, C. D. & Bailes, G. (2001). Assessing violence risk in mentally and personality disordered individuals. In C. R. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 71-84). Chichester: Wiley.
- Weichold, K. (2004). Evaluation eines Anti-Aggressivitäts-Trainings bei antisozialen Jugendlichen. *Gruppendynamik und Organisationsberatung*, 35, 83-104.
- Weidner, J. (1990). *Anti-Aggressivitäts-Training für Gewalttäter: Ein delikt spezifisches Behandlungsangebot im Jugendvollzug*. Bonn: Forum Verlag Godesberg.
- Weidner, J. & Wolters, J. M. (1991). Aggression und Delinquenz: Ein spezialpräventives Training für gewalttätige Wiederholungstäter. *Monatsschrift für Kriminologie und Strafrechtsreform*, 74, 210-223.
- Weisburd, D., Lum, C. M. & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science*, 578, 50-70.
- Weiss, C. H. (1997). Theory-based evaluation: Past, present, and future. *New Directions for Program Evaluation*, 76, 41-55.
- Wilson, D. B. & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6, 413-429.
- Wirth, W. (1996). Das Evaluierungskriterium der Legalbewährung: Ein methodologischer Problemaufriss in vier Thesen. In H.-J. Kerner, G. Dolde & H.-G. Mey (Hrsg.), *Jugendstrafvollzug und Bewährung: Analysen zum Vollzugsverlauf und zur Rückfallentwicklung* (S. 97-113). Mönchengladbach: Forum Verlag Godesberg.
- Wirth, W. (1998). Ausländische Gefangene im Jugendstrafvollzug NRW: Ergebnisse einer Stichtagserhebung (15. Juli 1997). *Zeitschrift für Strafvollzug und Straffälligenhilfe*, 47, 278-286.

- Wirth, W. (2002). Berufliche Bildung im Jugendstrafvollzug. In P. Fülbier & R. Münchmeier (Hrsg.), *Handbuch Jugendsozialarbeit: Geschichte, Grundlagen, Konzepte, Handlungsfelder, Organisation* (S. 768-774). Münster: Votum.
- Wirth, W. (2006a). Arbeitslose Haftentlassene: Multiple Problemlagen und vernetzte Wiedereingliederungshilfen. *Bewährungshilfe*, 53, 137-152.
- Wirth, W. (2006b). Ausbildungs- und Beschäftigungsintegration für Haftentlassene: die Entwicklungspartnerschaft MABiS.NeT. In A. Dessecker (Hrsg.), *Jugendarbeitslosigkeit und Kriminalität* (S. 257-275). Wiesbaden: Kriminologische Zentralstelle.
- Wolters, J.-M. (1992). *Kampfkunst als Therapie: Die sozialpädagogische Relevanz asiatischer Kampfsportarten*. Bern: Lang.
- Zamble, E. & Porporino, F. (1990). Coping, imprisonment, and rehabilitation: Some data and their implications. *Criminal Justice and Behavior*, 17, 53-70.
- Zeisel, H. (1982a). Disagreement over the evaluation of a controlled experiment. *American Journal of Sociology*, 88, 378-389.
- Zeisel, H. (1982b). Hans Zeisel concludes the debate. *American Journal of Sociology*, 88, 394-396.

Teile der kumulativen Dissertation

- (1) Bosold, C., Hosser, D. & Lauterbach, O. (2007). Psychosoziale Behandlung im Jugendstrafvollzug: Erste Ergebnisse einer Längsschnittstudie. *Praxis der Rechtspsychologie*, 17, 265-284.
- (2) Bosold, C. & Lauterbach, O. (2010). Leben ohne Gewalt organisieren: Evaluation eines Trainings für Gewalttäter im Jugendstrafvollzug. *Forensische Psychiatrie, Psychologie, Kriminologie*, 4, 267-277.
- (3) Hosser, D., Bosold, C. & Lauterbach, O. (2006). Sozialtherapeutische Behandlung von jungen Sexualstraftätern: Ergebnisse einer Evaluationsstudie. *Recht & Psychiatrie*, 24, 125-133.
- (4) Hosser, D., Lauterbach, O. & Camehn, K. (2008). Validität und Reliabilität des FPI-R beim Einsatz im Strafvollzug. *Diagnostica*, 54, 129-137.
- (5) Hosser, D., Lauterbach, O. & Höynck, T. (2007). Und was kommt danach? Entlassungsvorbereitung und Nachentlassungssituation junger Straftentlassener. In J. Goerdeler & P. Walkenhorst (Hrsg.), *Jugendstrafvollzug in Deutschland. Neue Gesetze, neue Strukturen, neue Praxis? Schriftenreihe der DVJJ Bd. 40* (S. 396-412). Mönchengladbach: Forum Verlag Godesberg.
- (6) Lauterbach, O. (2009). Soziale Integration und Delinquenz nach Entlassung aus dem Jugendstrafvollzug. *Zeitschrift für Jugendkriminalrecht und Jugendhilfe*, 20, 44-50.
- (7) Lauterbach, O. & Hosser, D. (2007). Assessing empathy in prisoners: A shortened version of the Interpersonal Reactivity Index. *Swiss Journal of Psychology*, 66, 91-101.