

**Sexueller Missbrauch von Minderjährigen durch Kleriker
und Laien im Verantwortungsbereich der Diözese Trier:
Eine ökosystemische Analyse aus den Perspektiven von
unmittelbar und mittelbar Betroffenen**

**Evaluation von Prognosegutachten
über Täter im Bistum Trier**

Petra Hank und Michelle Lange
Universität Trier, FBI – Psychologie
Trier, im September 2026

I. Evaluation von Prognosegutachten über die Täter

1. Einleitung

Psychiatrisch-psychologische Begutachtungen sind von zentraler Bedeutung für die prognostische Einschätzung des Rückfallrisikos pädosexueller Täter. Im Rahmen der vorliegenden Untersuchung fungierten entsprechende Prognosegutachten, welche seitens des Bistums bezüglich einzelner Täter in Auftrag gegeben worden waren, als Datenbasis. Die inhaltsanalytische Auswertung dieser Dokumente zielte darauf ab, handlungsleitende Motive sowie Täterstrategien zu identifizieren und die Perspektive der Beschuldigten auf das Missbrauchsgeschehen zu rekonstruieren. Eine tragfähige Datengrundlage setzt voraus, dass die Gutachteninhalte valide und verlässlich sind. Um die Aussagekraft des vorliegenden Datenmaterials zu überprüfen, wurde eine kriteriengeleitete Evaluation der Prognosegutachten durchgeführt, deren Ergebnisse nachfolgend dargelegt werden. Dem stellt sich eine Erläuterung der Funktionen psychiatrisch-psychologischer Prognosegutachten sowie eine Deskription der wissenschaftlichen Qualitätsanforderungen an den Begutachtungsprozess voran.

2. Theoretischer Hintergrund

2.1. Funktionen psychiatrisch-psychologischer Prognosegutachten

Ein psychiatrisch-psychologisches Prognosegutachten, ab hier Gutachten, ist mit einer Reihe von Erwartungen verbunden. In erster Linie dient es einer Rückfallprognose: Es soll eine fundierte Einschätzung des Risikos ermöglichen, dass sich ein bereits wegen sexuellem Kindesmissbrauch auffällig gewordener Beschuldigter künftig erneut gegenüber Minderjährigen sexuell übergriffig verhalten wird. Damit sollte das Gutachten potenzielle Opfer vor weiteren möglichen Straftaten schützen (Rettenberger, 2017). Das Gutachten sollte auch erklären, warum dieser Beschuldigte zu diesem Zeitpunkt einem bestimmten Kind /Jugendlichen sexuelle Gewalt angetan hat (Stichwort: idiographisches Erklärungsmodell; Dahle, 2005). Über dies hinaus sollte die Begutachtung zum Risikomanagement beitragen, indem sie Anhaltspunkte für die Planung und Durchführung von (präventiven) Interventionsmöglichkeiten (z. B. Kontrollmaßnahmen, Führungsaufsicht, Therapie) aufzeigt, die das Rückfallrisiko verringern (Dahle, 2005; Rettenberger, 2017). Zudem erwarten Boetticher et al. (2007) das

Aufzeigen von hypothetischen Rückfallszenarien. So gesehen sollte das Gutachten auch dazu dienen, den Beschuldigten zu befähigen, Risikosituationen zu vermeiden bzw. besser mit ihnen umzugehen (Stiels-Glenn, 2017). Zusammengenommen schafft ein Gutachten die Voraussetzungen, auf deren Grundlage der Auftraggeber die Frage, ob ein Beschuldigter erneut Minderjährige sexuell missbrauchen wird, in eigener Wertung und Würdigung beantworten und in eigener Verantwortung eine Entscheidung treffen kann (z. B. Dreßing & Foerster, 2021; Egli-Alge, 2017). Damit sichert das Gutachten den Entscheidungsträger nach außen ab. Und schließlich werden durch die gutachterliche Tätigkeit wissenschaftliche Erkenntnisse in die Öffentlichkeit vermittelt (Dreßing & Foerster, 2021).

Zusammengenommen beinhalten die Funktionen der Gutachten unterschiedliche Erwartungshaltungen mit zum Teil konträren Zielvorstellungen. Sie verdeutlichen nicht nur das „mehrfach determinierte Verantwortungsverhältnis“ der gutachtenden Person (Foerster, 2003), sondern machen auch auf die ethische Dimension der gutachtlichen Tätigkeit aufmerksam: Das Gutachten ist zugleich eine Handlungsempfehlung (Steinböck, 2017); sie greift in das zukünftige Leben der begutachteten Person ein und gestaltet es mit (z. B. Gretenkord, 2017; Stiels-Glenn, 2017; über eine ethisch angemessene gutachterliche Tätigkeit informieren z. B. die Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e. V. und der Deutschen Gesellschaft für Psychologie e. V. (2022), die der American Academy of Psychiatry and the Law (2005) und die Speciality Guidelines for Forensic Psychology der American Psychological Association (2012)).

2.2. Qualitätsanforderungen an psychiatrisch-psychologische Prognosegutachten

Angesichts ihrer entscheidenden Bedeutung für die begutachtete Person werden hohe Qualitätsanforderungen an die Gutachten gestellt (z. B. Boetticher et al., 2007, 2019; Kröber et al., 2019). Dieser Qualitätsanspruch wurde oftmals in der prognostischen Begutachtungspraxis nicht zufriedenstellend erfüllt (zusammenfassend Wertz et al., 2018, 2023). Die entsprechenden Studien prüften mehrheitlich die Qualität von Gutachten zu Schuldfähigkeit und Kriminalprognose. Dabei stand die Durchführung der psychiatrisch-psychologischen Untersuchung an sich und die Darstellung der Untersuchungsergebnisse im Fokus. Beurteilungsgrundlage waren allgemeine Mindestanforderungen mit geringem Bezug zu empirischer Evidenz (Kindler & Fegert, 2023).

Festgestellt wurden zum Teil defizitäre Anamnesen (z. B. Nowara, 1995), oftmals unangemessene Untersuchungszeiten (z. B. Kury & Adams, 2010), teilweise mangelhaft dokumentierte Untersuchungsergebnisse (z. B. Prüter-Schwarte et al., 2019;) und mitunter unzureichende Methodenkenntnisse der Sachverständigen (z. B. Nedopil, 2005). Über die Methodik hinaus wurden in einigen Arbeiten (z. B. Galli, 2011; Prüter-Schwarte et al., 2019; Wertz et al., 2018) die wissenschaftlichen Grundlagen der Gutachten untersucht. Auch hierzu fielen die Ergebnisse in Teilen unbefriedigend aus. Über die Qualität von psychiatrisch-psychologischen Gutachten außerhalb der strafrechtsrelevanten Begutachtungspraxis liegen bisher nur wenige empirische Belege vor. So befindet sich bspw. die Forschung zur Qualität der Begutachtungspraxis im Familienrecht „in einem frühen Stadium“ (Kindler & Fegert, 2023, S. 589). Erste Befunde hierzu offenbarten gravierende Qualitätsmängel für den überwiegenden Teil der untersuchten Gutachten (Salewski & Stürmer, 2014).

Zusammenfassend ist festzuhalten, dass die empirischen Studien die psychiatrische und psychologische Gutachtenpraxis im Gerichtswesen evaluierten. Die Ergebnisse dieser wissenschaftlichen Untersuchungen belegen, dass Fehlerquellen in der psychiatrischen und psychologischen Gutachtenpraxis im Gerichtswesen ein ernst zu nehmendes Problem darstellen. Angesichts der positiven Korrelation zwischen der Qualität der Gutachten und ihrer prognostischen Gültigkeit wiegen diese Ergebnisse schwer (Wertz et al., 2018, 2023). Zugleich gibt es eine große Anzahl an empirischen Forschungsarbeiten, die nachweist, dass Spezifika der Begutachtungsgegenstände und des Begutachtungsprozesses mögliche Fehlerquellen im Begutachtungsprozess beeinflussen können (Oberlader & Schmidt, 2024). Vor diesem Hintergrund leitet sich die wichtige Frage nach der Qualität der Begutachtungspraxis im Kontext der Begutachtung der pädosexuellen Kleriker im Bistum Trier ab, einem Begutachtungskontext *außerhalb* von juristischen Anwendungsfeldern, wo die Verwertbarkeit der Gutachten durch Gerichte irrelevant ist, die Verständigung über die Qualität eines Gutachtens möglicherweise von den Interessen des Auftraggebers geleitet wird und wo keine sachkundige Qualitätskontrolle stattfindet.

Ziel der vorliegenden Arbeit war es daher, die wissenschaftliche Herangehensweise in der psychiatrisch-psychologischen Begutachtungspraxis der pädosexuellen Kleriker im Bistum Trier zu untersuchen. Die Analyse sollte Aufschluss geben über die Transparenz, Nachvollziehbarkeit und Überprüfbarkeit in dem Vorgehen der

Gutachter und über die wissenschaftliche Begründung der Gutachten informieren. Weiterführend sollten die Ergebnisse die Frage beantworten, ob die vorliegend analysierten Gutachten als Datenquellen für eine inhaltsanalytische Rekonstruktion des Missbrauchsgeschehens aus Beschuldigtenperspektive taugen. Die gutachterliche Tätigkeit wurde im Sinne einer entscheidungsorientierten Einzelfalldiagnostik aufgefasst (Sterner et al., 2024), die in Handlungsempfehlungen unter Risiko über die untersuchte Person mündet (Cronbach & Gleser, 1965).

2.3 Der psychiatrisch-psychologische Prognoseprozess: Verfahrensablauf und Qualitätsstandards

Verschiedene Handlungsmodelle strukturieren den Prozess der psychiatrisch-psychologischen Prognosestellung (z. B. Dahle, 2005; Dreßing & Foerster, 2021; von Franqué, 2013; Prüter-Schwarte, 2022; Rettenberger, 2017). Die Modelle unterscheiden sich in der Anzahl und in der Differenziertheit der beschriebenen Prozessschritte (s. hierzu zsf. von Franqué, 2013). Im Kontext der forensisch-psychiatrischen Begutachtungen empfehlen bspw. Dreßing & Foerster (2021) ein dreistufiges Vorgehen, beginnend mit der Feststellung der psychopathologischen Diagnose, gefolgt von deren Übersetzung in juristische Fachtermini und der Erörterung und Beantwortung der Beweisfragen im dritten Schritt. Unabhängig vom jeweiligen Rechtsbereich grenzen Segmiller und Dudeck (2019) die Datenerhebung, die Befundlegung und die Urteilsbildung als Aufgaben einer psychiatrischen Begutachtung voneinander ab. Dahle (2005) will zusätzlich die Frage nach der Indikation von geeigneten Interventionsmöglichkeiten beantwortet wissen, um Risikofaktoren der Rückfälligkeit zu minimieren. Angelehnt an Prozessmodelle der klinisch-psychologischen Diagnostik beschreibt von Franqué (2013) den Ablauf der kriminalprognostischen Urteilsbildung in acht Teilschritten. Westhoff und Kluck (2003) gliedern den Prozess der gutachterlichen Tätigkeit im Kontext der psychologischen Begutachtung in neun Schritte; diese sind: Definition des Gutachtenauftrags, Darlegung der Gesamtfragestellung, Ableitung der Teilfragestellungen, Untersuchungsplanung, Untersuchungsdurchführung, Ergebnisbericht, Befund, Beantwortung der Gesamtfragestellung, Schlussfolgerungen. Vorliegend diene dieses Prozessmodell als Bewertungsgrundlage, da es eine hochdifferenzierte Auflösung der gutachterlichen Tätigkeit bietet, die alle inkludierten Prozessschritte in ihrer wechselseitigen Beziehung erfasst. Ihre Analyse ermöglicht, den Prozess der Prognosebegutachtung grundlegend zu durchleuchten, mögliche

Schwachstellen in diesem Prozess offen zu legen und Ansatzpunkte zu seiner Optimierung aufzuzeigen.

In einer Literaturanalyse wurden die relevanten Anforderungen an die Schritte des gutachtlichen Prozesses herausgearbeitet. Da der Qualitätsbegriff und die daraus abzuleitenden Anforderungen, die an ein Gutachten gestellt werden, zwischen den Fachdisziplinen divergieren (Kindler & Fegert, 2023), wurde das Qualitätsverständnis von Experten und Expertinnen unterschiedlicher Professionen berücksichtigt. Als Grundlage wurden herangezogen die Arbeiten von Boetticher et al. (2007, 2019), Dahle und Lehmann (2018), Fegert et al. (2006), Graumann (1966), Kaminski (1970), Kane (2013), Kröber et al. (2019), Nedopil (2022), Rettenberger und Brettel (2020), Rettenberger und von Franqué (2013), Salewski und Stürmer (2014), Sterner et al. (2024), Tack (1976), Westhoff und Kluck (2013) sowie die Empfehlungen des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen (2017). Die Anforderungen werden nachfolgend in Bezug auf die einzelnen Prozessschritte (PS) des Begutachtungsprozesses dargestellt.

Der *Gutachtenauftrag* sollte die Fragestellung des Auftraggebers allgemeinverständlich definieren (PS 1) und eindeutig zum Ausdruck bringen, welche Entscheidungsgrundlage durch das Gutachten erarbeitet werden soll. Da die allgemeinsprachliche Formulierung des Gutachtenauftrags nicht den Anforderungen an die Wissenschaftssprache genügt, sind die Fragestellungen des Auftraggebers in eine, semantisch eindeutige, Fachterminologie zu „übersetzen“. Damit einhergehend sind die erforderlichen diagnostischen Strategien und Zielsetzungen zu antizipieren (PS 2). Aus der Gesamtfragestellung werden psychiatrische und psychologische *Teilfragen* abgeleitet. Sie strukturieren den weiteren Prozess und entsprechen den zu prüfenden Hypothesen (PS 3). Grundlage einer empirisch informierten Hypothesenbildung sind Bezüge auf die Fachliteratur. Vorliegend sind dies Rekurse auf theoretische Annahmen und empirische Befunde, u. a. zu Merkmalen von Sexualtätern (Alter zum Zeitpunkt der ersten Tat, Hinweise auf psychische Auffälligkeiten, Berücksichtigung von relevanten Komorbiditäten, Delinquenz, Einstellung zum Tatgeschehen) sowie zu Risiko- und Schutzfaktoren ihrer Rückfälligkeit (Schwere und Häufigkeit der Taten, Gefühle von Scham und Schuld, Empathie für das Opfer; s. zsf. dazu Kröber et al., 2019; Kurtz, 2017; Rettenberger, 2019; Yoon et al., 2018). Das Risk-Need-Responsivity-Modell (RNR; Bonta & Andrews, 2017), das integrative Modell für die

Kriminalprognose bei pädosexuellen Delikten (Rettenberger, 2017) oder auch die Klassifikation von Risikofaktoren nach von Franqué (2013) stellen das notwendige Wissen für eine Analyse des individuellen Bedingungsgefüges personeller und situativer tatassoziierten Faktoren zur Verfügung. Darauf aufbauend sind normative Strategien zu entwickeln. Sie regeln das Entscheidungsverhalten des Gutachters bei gegebenen Daten und Zielsetzungen. Dazu sollten die Teilfragen Prädiktor-Kriteriums-Beziehungen explizieren und die individuelle Bedeutsamkeit der Prädiktoren bzw. ihrer Kombination für die Rückfälligkeit herausarbeiten. Die relevanten Konstrukte sind im Zuge der *Untersuchungsplanung* (PS 4) multimethodal zu operationalisieren. Der Untersuchungsplan sollte über die diagnostischen Zielsetzungen der eingesetzten Verfahren informieren. Jedes Verfahren (Dokumentenanalyse, Anamnese, klinisches Interview, Verhaltensbeobachtung und psychometrische Testverfahren; für eine Systematik der Verfahren s. Rettenberger, 2018) ist so zu beschreiben, dass eine Prüfung seiner Zulänglichkeit möglich ist. Die Ausprägungen der Indikatorvariablen sind regelgeleitet zu spezifizieren, so dass ein Anforderungsprofil resultiert. Die Schritte PS 1 bis PS 4 lassen sich zu dem Prozessabschnitt (PA) *Konzeptualisierung der gutachtlichen Tätigkeit und Planung der Datenbeschaffung* zusammenfassen.

Die psychiatrisch-psychologische *Untersuchung* (PS 5) sollte mit einer informierten Einwilligung der, zu begutachtenden, Person unter standardisierten Bedingungen durchgeführt werden. Sie ist so zu dokumentieren, dass ihre intersubjektive Nachprüfbarkeit und Replikation möglich sind. Der *Ergebnisbericht* (PS 6) sollte alle erhobenen Ergebnisse, getrennt nach den Untersuchungsmethoden, korrekt und nachvollziehbar beschreiben. Reliabilität, Validität und Auswertungsobjektivität der Ergebnisse sollten gesichert sein. Zusammenfassend beschreiben PS 5 und PS 6 die Phase der *Datenerhebung, -auswertung und Ergebnisdarstellung*.

Es folgt der Prozessabschnitt *Interpretation und Integration der Daten, Empfehlung von Maßnahmen* (PS 7 bis PS 9). Der *Befund* (PS 7) beantwortet die Teilfragen im Einzelnen, indem die zugehörigen Ergebnisse interpretiert werden; theoretische Aussagen und empirisch gesicherte Erkenntnisse sollten in die Beurteilung der Befunde einbezogen werden. Psychopathologische Symptome sind in ihren Auswirkungen auf die Handlungs- bzw. Steuerungsfähigkeit der begutachteten Person darzustellen. Die gutachterliche Tätigkeit schießt mit der *Beantwortung der Fragestellung* (PS 8) und einer entscheidungsorientierten *Ableitung von Handlungsempfehlungen*

gen (PS 9). Die Einzelbefunde werden zu einem evidenzbasierten Urteil integriert. Dabei sind individuelle Risiko- und Schutzfaktoren der Rückfälligkeit zu berücksichtigen. Es ist eine Empfehlung für die Handlungsalternative zu geben, die, basierend auf den Untersuchungsergebnissen, das Entscheidungskriterium optimiert und die erwarteten Kosten minimiert.

2.4. Fragestellungen

Ziel der vorliegenden Arbeit war es zum einen, die psychiatrisch-psychologische Begutachtungspraxis zur Prognose der Rückfälligkeit von pädosexuellen Klerikern im Bistum Trier zu untersuchen. Es sollten die Algorithmen und Heuristiken identifiziert werden, die der Gutachter für die Risikoeinschätzung anwendet. Ob diese inhaltlich *richtig* ist, war nicht Gegenstand der vorliegenden Arbeit. Das Ergebnis dieser methodischen Analyse sollte weiterführend einen Anhaltspunkt dafür liefern, ob auf Grundlage der Gutachten eine valide inhaltsanalytische Rekonstruktion des sexuellen Missbrauchs aus der Perspektive der Kleriker möglich ist. Ausgehend von diesen Zielsetzungen wurde vorliegend eine kriteriengeleitete Evaluation des gesamten Begutachtungsprozesses durchgeführt. Im Einzelnen wurden die Prozessabschnitte *Konzeptualisierung der gutachtlichen Tätigkeit und Planung der Datenbeschaffung* (PS 1 bis PS 4), *Datenerhebung, -auswertung und Ergebnisdarstellung* (PS 5, PS 6) sowie *Interpretation und Integration der Daten, Empfehlung von Maßnahmen* (PS 7 bis PS 9) bewertet. Im Einzelnen wurden folgende Fragestellungen (F) geprüft; sie entsprechen den Schritten im Begutachtungsprozess.

- F1 Definiert der Gutachtauftrag die Fragestellung des Auftraggebers eindeutig, so dass klar und verständlich ist, welche Entscheidungsgrundlage durch die Begutachtung erreicht werden soll?
- F2 Wird die Fragestellung des Auftraggebers korrekt in psychiatrische bzw. psychologische Fachtermini übersetzt?
- F3 Sind die zu prüfenden Teilfragen stringent aus einem rational begründeten Gedankengebäude abgeleitet, das dem aktuellen Stand der Wissenschaftsdisziplin entspricht?
- F4 Erfolgt die Durchführung der diagnostischen Untersuchung planvoll?
- F5 Wird die Durchführung der diagnostischen Untersuchung intersubjektiv nachprüfbar dokumentiert?

- F6 Berichtet die Ergebnisdarstellung die, einer Untersuchungsmethode zugehörigen, diagnostischen Informationen jeweils vollständig und korrekt?
- F7 Verdichtet die Befundung die, einer Teilfrage zugehörigen Ergebnisse jeweils zu einer wissenschaftlich begründeten und nachvollziehbaren Aussage?
- F8 Beantwortet die Integration der Einzelbefunde die Gesamtfragestellung evidenzbasiert und schlüssig?
- F9 Werden ausgewogene und logisch nachvollziehbare Handlungsoptionen empfohlen?

3. Methode

Für die Analyse des gutachterlichen Vorgehens wurde das *Evaluationsinstrument für Gutachten (EvaGut)* entwickelt. Angelehnt an das Modell von Westhoff und Kluck (2013) bildet EvaGut den gutachterlichen Prozess ab. Es gliedert sich in die zuvor genannten Abschnitte des Begutachtungsprozesses mit den zugehörigen Schritten. Ausgehend von den beschriebenen Anforderungen wurden für die Evaluation der Prozessschritte jeweils zwei bis zehn inhaltliche Bewertungskriterien bestimmt und in Form von Aussagen, den Prüfitems, operationalisiert. Anhang I 1 informiert über die Inhalte und Auswertung der Kriterien von EvaGut, die Wertebereiche der Kriterienscores sowie deren Interpretation (siehe hierzu die Anhänge I 2 und I 3).

3.1. Stichprobe und Durchführung

In die Analyse wurden 23 Gutachten über pädosexuelle Kleriker, in Auftrag gegeben durch das Bistum Trier, einbezogen. Sie wurden erstellt im Zeitraum von 2006 bis 2020 (auf die Jahre 2006, 2008 und 2009 entfiel jeweils ein Gutachten, in den Jahren 2010, 2011, 2012, 2013, 2014 und 2020 waren es acht, zwei, fünf, ein, drei und ein Gutachten). Alle analysierten Gutachten wurden mit einer Ausnahme nach der Veröffentlichung der „Mindestanforderungen an Prognosegutachten“ (Boetticher et al. 2007) erstattet. Die Auftragsvergabe erfolgte nach Auskunft des Bischofs aufgrund von Empfehlungen. 22 der Gutachten fertigte ein und derselbe Psychiater in Zusammenarbeit mit einem Psychologischen Psychotherapeuten an, wobei in 21 der 22 Gutachten derselbe Kollege mitarbeitete. Ein Gutachten erstattete ein Psychiater

ohne die Mitarbeit eines weiteren Kollegen. Alle Gutachter waren männlichen Geschlechts. Anlässe der analysierten Gutachten waren sexuelle Übergriffe an Minderjährigen, begangen von Priestern im Bistum Trier. Die Tatzeitpunkte reichten von 2019 bis in die 1960er Jahre zurück. In 21 Gutachten (91.3 %) wurde die Bearbeitungsdauer berichtet. Sie betrug durchschnittlich $M = 3.4$ Monate ($SD \pm 1.1$ Monate) und $M = 12.8$ Tage ($SD \pm 9$ Tage) laut angegebenem Auftrags- und Erstattungsdatum. In 21 Gutachten (91.3 %) war die Gesamtuntersuchungszeit dokumentiert. Wurde ausschließlich eine Exploration durchgeführt (38 %), lag die Untersuchungszeit im Durchschnitt bei $M = 3.2$ Stunden ($SD \pm 0.37$ Stunden). Kam eine testpsychologische Untersuchung hinzu (62 %), verlängerte sich die Gesamtuntersuchungszeit auf durchschnittlich $M = 4.3$ Stunden ($SD \pm 1.2$ Stunden). Der Umfang der Gutachten variierte zwischen 16 und 31 Seiten ($M = 20.3$ Seiten; $SD \pm 4$ Seiten).

Die Autorinnen evaluierten alle Gutachten unabhängig voneinander mit Hilfe von EvaGut. Zur Überprüfung der Inter-Rater-Reliabilität (IRR) wurde mit SPSS, Version 29.0.2.0, für Krippendorffs α (Krippendorff, 1970) ein Wert von .96 (95%-Konfidenzintervall: .89 – 1, basierend auf 10.000 Bootstrap-Samples), gemittelt über die Bewertung der einzelnen Prozessschritte, berechnet. Dieser Wert weist auf eine sehr hohe Übereinstimmung zwischen den Beurteilungen hin (Krippendorff, 2019).

4. Ergebnisse

4.1. Konzeptualisierung und Planung der Datenbeschaffung

Im Mittel konstituierten vier Fragestellungen den Gutachtenauftrag (Minimum = 1 Frage, Maximum = 7 Fragen). Sie betrafen die Frage nach der Störungsdiagnose, der Einschätzung der Rückfallwahrscheinlichkeit, der Zuweisung zu einem Arbeitsfeld und die Frage nach der Indikation einer therapeutischen Maßnahme. Die Fragen waren allgemeinsprachlich und verständlich formuliert. (PS 1-Bewertung: 2 von 2 Punkten, in Worten: sehr gut). Die erste Forschungsfrage ist mit „ja“ zu beantworten.

In keinem der analysierten Gutachten wurden die Fragestellungen in Fachtermini übersetzt noch in Bezug zu den damit verbundenen diagnostischen Strategien gebracht (PS 2-Bewertung: 0 Punkte, in Worten: mangelhaft). F2 ist mit „nein“ zu beantworten.

Kein Gutachten leitete aus den Fragestellungen des Auftraggebers psychiatrische und / oder psychologische Teilfragen ab. Es wurden auch keine Anforderungen an die Ausprägungen der zu untersuchenden tatassozierten Faktoren (Prädiktoren) bzw. ihre Konstellation expliziert. Folglich ermangelten alle Prognosegutachten ein Anforderungsprofil. Für alle analysierten Gutachten bewerteten die Autorinnen diesen Prozessschritt übereinstimmend als „mangelhaft“ (0 Punkte); F3 ist mit „nein“ zu beantworten.

In Bezug auf die Planung der diagnostischen Untersuchung war für alle Gutachten eine Fehlanzeige festzustellen: Es wurden keine Angaben dazu gemacht, welche tatassozierten Faktoren wie operationalisiert und wozu erhoben werden sollten (PS 4-Bewertung: 0 Punkte, in Worten: mangelhaft); F4 ist mit „nein“ zu beantworten.

4.2. Datenerhebung, -auswertung und Ergebnisdarstellung

Ort, Zeit und Dauer der diagnostischen Untersuchung wurden angegeben, die Gutachtenden benannt. Wer von den beiden Gutachtenden welchen Part der Untersuchung übernahm, wurde nicht dokumentiert. In allen Gutachten wurde auf die informierte Einwilligung der Begutachteten hingewiesen. Die Datenbeschaffung war idio-graphisch orientiert und erfolgte multimethodal (s. Anhang I 4). In allen Begutachtungen wurden mündliche Befragungen (Anamnese, Exploration, Interview) und Verhaltensbeobachtungen durchgeführt. Es wurde multivariate Information erhoben. Durchgeführt wurden jeweils eine biographische Anamnese mit Angaben zur biologischen Entwicklung, zur Herkunftsfamilie und dem Lebenslauf, eine Sexualanamnese, eine medizinisch-psychiatrische Anamnese, eine Suchtmittelanamnese sowie eine Exploration zum Anlass der Untersuchung und zu tatassozierten Merkmalen. Für diese Verfahren fehlten Hinweise auf Instruktionen zu ihrer Durchführung wie auch Regeln zur Auswertung der erhobenen Daten. Im Rahmen der mündlichen Befragungen wurde außerdem das Verhalten der Begutachteten beobachtet. Keine gutachtliche Tätigkeit erwähnte Hinweise auf eine standardisierte Verhaltensbeobachtung. In 14 der 23 Begutachtungen wurden zusätzlich bis zu zwei multidimensionale Persönlichkeitsinventare verwendet. Zum Einsatz kamen in 14 Gutachten der Gießen-Test (GT; Beckmann, et al., 1991), in zwei Gutachten die deutschsprachige Version des Minnesota Multiphasic Personality Inventory (MMPI-2; Engel et al., 2000). In keinem Gutachten wurde eine Indikation für ihre Verwendung angegeben; Aufbau und Messin-

tention der Verfahren wurden im Rahmen der Ergebnisdarstellung unvollständig beschrieben. Überdies wurden Sachinformationen aus Dokumenten (z. B. Arztbriefe, Gesprächsprotokolle) erschlossen, die vom Auftraggeber zur Verfügung gestellt wurden. In 22 Gutachten wurde auf einzelne dieser Dokumente verwiesen; ob die Dokumentenanalyse etwaig darüber hinausging, war nicht eindeutig feststellbar. Lediglich in einem Gutachten wurden alle zur Verfügung gestellten Unterlagen aufgelistet. In welcher Reihenfolge die Daten erhoben wurden, ging nicht aus den Gutachten hervor. Zusammengefasst befanden die Autorinnen die Qualität der Durchführung der psychiatrisch-psychologischen Untersuchung als ausreichend (PS 5-Bewertung: 3 Punkte); F5 ist mit „nein“ zu beantworten.

Die Ergebnisse der mündlichen Befragungen wurden jeweils anhand von protokollartigen, nach Themenfeldern getrennten, Gesprächsverläufen berichtet. Da Hinweise zur Aufzeichnung und Auswertung der Gesprächsdaten fehlten, entzogen sich die Angemessenheit ihrer Auswertung und die Vollständigkeit ihrer Darstellung einer Prüfung. Mit einer Ausnahme wurden in allen Gutachten Verhaltensbewertungen referiert. Sie basierten auf beobachtetem Verhalten mit einem konkreten zeitlichen und Handlungsbezug. Aufgrund fehlender Hinweise zur Aufzeichnung und Auswertung der Beobachtungsdaten, waren die Korrektheit der beobachtungsbezogenen Ergebnisse und die Vollständigkeit ihrer Darstellung nicht prüfbar. Psychologische Störungen wurden anhand der Klassifikationssysteme ICD-10 bzw. DSM-IV gestellt; Auswirkungen der psychischen Beeinträchtigungen auf die Einsichts- und Handlungsfähigkeit der Begutachteten wurden nicht dargestellt. In Bezug auf die Ergebnisdarstellung der psychologischen Testdiagnostik waren fehlende Hinweise auf die Auswertung der Verfahren festzustellen. Ob die Daten verfahrenskonform ausgewertet wurden, ließ sich daher nicht beurteilen. Ebenso fehlten Angaben zu den genutzten Normen, so dass die Angemessenheit der jeweiligen Referenzstichproben nicht geprüft werden konnte. Es wurde über die Testergebnisse informiert, indem die Testwerte aus den Verfahren direkt mit den Normwerten verglichen wurden. Testwerte, die zwischen plus und minus einer Standardabweichung um den Mittelwert in der Normstichprobe lagen, wurden als durchschnittlich bewertet. Die Fehlerbehaftung der Testwerte blieb dabei unberücksichtigt. Darüber hinaus wurden keine Entscheidungsgrundlagen zur Beurteilung der Fehlerrisiken genannt. Ob sich die Gutachter von der üblichen Konvention leiten ließen und jeweils Konfidenzintervalle mit einem Signifikanzniveau von $\alpha = 0.05$ angenommen wurden, blieb unklar. Weiter war festzustel-

len, dass in keinem Gutachten vollständig über die Ergebnisse für alle Dimensionen der Testverfahren informiert wurde. Die gesamte Ergebnisdarstellung bewerteten die Autorinnen übereinstimmend als mangelhaft (PS 6-Bewertung: 0 Punkte); F6 ist mit „nein“ zu beantworten.

4.3. Interpretation und Integration der Daten, Empfehlung von Maßnahmen

In jedem Gutachten wurden im Begutachtungsschritt „Beurteilung“ alle Teil-Fragestellungen des Gutachtauftrags mittels klinischer Urteilsbildung (Meehl, 2013) beantwortet, mutmaßlich basierend auf der Expertise der Gutachtenden. Individuelle Risikofaktoren blieben dabei außen vor. In allen Gutachten wurde auf eine durchgängige theoretische und empirische Einordnung der diagnostischen Informationen verzichtet. In 16 Gutachten (70 %) wurden Limitationen der Begutachtung diskutiert (PS 7 / PS 8-Bewertung: 4 bis 6 Punkte, in Worten: ausreichend für 83 %, mangelhaft: 17 %). F7 und F8 sind mit „nein“ zu beantworten. In 22 Gutachten wurden zwei Handlungsempfehlungen formuliert, Bedingungen zu ihren Verwirklichungen wurden in 18 Gutachten aufgezeigt. Die Ableitung der Maßnahmen gründete maßgeblich auf dem (Nicht-)Vorliegen einer Störungsdiagnose. In keinem Gutachten wurden mögliche Fehlerrisiken beurteilt, das Kostenverhältnis der Handlungsoptionen wurde nicht dargelegt (PS 9-Bewertung: 2 bis 4 Punkte, in Worten: mangelhaft für 17 %, ausreichend für 53 %, befriedigend: 17 %, gut: 13 %). F9 ist für ein Teil der Prognosestellungen mit „ja“ zu beantworten.

Zusammenfassend war für die vorliegend analysierten Gutachten eine mangelhafte Qualität in Bezug auf die Begutachtungsabschnitte, *Konzeptualisierung der gutachtlichen Tätigkeit und Planung der Datenbeschaffung*, und *Datenerhebung, -auswertung und Ergebnisdarstellung* festzustellen. Für den dritten Prozessabschnitt, *Interpretation und Integration der Daten, Empfehlung von Maßnahmen* ergab die Evaluation eine mangelhafte (31 %) bis ausreichende Qualität (69 %). Über den gesamten Prozess der psychiatrisch-psychologischen Begutachtung hinweg betrachtet, waren alle analysierten Gutachten von mangelhafter Qualität.

5. Diskussion

Im Ergebnis der kriteriengeleiteten Evaluation waren für alle vorliegend analysierten Gutachten erhebliche Defizite festzustellen. Sie umspannen den gesamten Begutachtungsprozess.

5.1. Konzeptualisierung und Planung

Die alltagssprachlich formulierten Fragestellungen des Auftraggebers sind Ausgangspunkt der psychiatrisch-psychologischen Untersuchungen. Dadurch mangelt es den Untersuchungen zum einen an einer operationalen Präzisierung ihrer Zielsetzungen, zum anderen bleiben sie mit dem, an die Alltagssprache geknüpften, Laien-Wissen verhaftet (Kaminski, 1970). Bereits an dieser Stelle verfehlen die Gutachten ihren wissenschaftlichen Anspruch. Des Weiteren wird die jeweilige Konzeption einer Begutachtung nicht ausgeführt. So wird die Relevanz von psychopathologischen Auffälligkeiten im Allgemeinen bzw. die von Störungen der Sexualpräferenz für das Ausüben pädosexuellen Verhaltens nicht erörtert. Komorbiditäten mit prognostischer Relevanz (Philipp & Rettenberger, 2025) werden ebenso wenig berücksichtigt. Weiter ist den Prognosegutachten nicht zu entnehmen, welche tatassozierten Merkmale theoriegeleitet als relevant für die zu bearbeitenden Fragestellungen erachtet werden, welche empirisch fundierten Annahmen zu deren weiterer Entwicklung in einem definierten Zeitraum erstellt werden und wie diese Merkmale zu empirisch begründeten Entscheidungsstrategien konfiguriert werden. Mit anderen Worten, es werden keine einzelfallbezogenen Konstellationen von Risiko- und Schutzfaktoren pädosexuellen Verhaltens dargelegt. Somit entzieht sich die wissenschaftliche Grundierung der Gutachten einer Beurteilung. Es bleibt offen, ob die Gutachtenden auf erfahrungswissenschaftlich fundiertes Wissen rekurrierten oder auf einschlägige Expertise, für die Mängel der Subjektivität (z. B. subjektive Selektion, Akzentuierung, Wertung, unangemessene Denkmodelle) nicht auszuschließen sind (Kaminski, 1970). Außerdem fehlen explizite normative Strategien, die das Entscheidungsverhalten der Gutachtenden bei gegebenen Daten und Zielsetzungen regeln und transparent machen (Tack, 1976). Somit fehlt ein kommuniziertes erfahrungswissenschaftliches Fundament, um im Einzelfall die tolerierbare falsch-positive bzw. falsch-negative Irrtumswahrscheinlichkeit begründet zu entscheiden (Sterner et al., 2024). Dadurch erhöhen sich der Interpretationsspielraum und die Entscheidungsunsicherheit im weite-

ren Begutachtungsprozess und machen erratische Entscheidungen wahrscheinlicher (Murrie & Warren 2005).

Auch mit Blick auf die Planung der psychiatrisch-psychologischen Untersuchung sind den analysierten Gutachten keine Hinweise zu entnehmen. Die strategische Planung unterlassen zu haben, bedeutete ohne diagnostische Strategien begutachtet zu haben. Damit fehlten Regeln, die angeben, unter welcher Zielsetzung (Störungsdiagnose, Prognose) bei gegebenen Informationen zu entscheiden wäre. Mithin fehlten Aussagen darüber, wie die erhobenen Daten benutzt werden sollten, und welche empirisch fundierten Annahmen zu Prädiktoren pädosexuellen Verhaltens die spezifischen Interpretationen rechtfertigten (Kane, 2013). Somit bliebe unklar, ob entscheidungsrelevante Prädiktoren evidenzbasiert für die diagnostische Untersuchung ausgewählt wurden. Dieses Versäumnis hätte Konsequenzen für die Auswahl der eingesetzten Erhebungsmethoden: Den durchgeführten Anamnesen, Explorationen und Interviews fehlte mit dem Auslassen der strategischen Datenplanung ein unverzichtbares Steuerungselement für deren Inhalt und Ablauf (Kaminski, 1970). Den Verhaltensbeobachtungen fehlte ein psychiatrisch-psychologisches Beobachtungssystem, das inhaltliche und formale Kriterien der begründetermaßen wichtigen Verhaltensmerkmale definierte. Diese gleichsam freien Beobachtungen leisteten einer Reihe von Beobachtungsfehlern Vorschub (Hock et al., 2023; Kanning, 2000; Kochinka, 2010). Die Grundlage einer wissenschaftlichen Beobachtung wäre vorliegend nicht gegeben (Grewe & Wentura, 1997). Für die eingesetzten testpsychologischen Verfahren (Gießen-Test, MMPI-2) hätte keine Prüfung ihrer „passenden Indikationsstellung“ (Steinböck, 2017) stattgefunden, so dass die inkrementelle Validität dieser Verfahren für die jeweilige Prognosestellung unklar bliebe. Für die eingesetzten testpsychologischen Verfahren wäre demnach ungewiss, ob die Verfahren mit der höchsten prädiktiven Validität eingesetzt wurden. Geht man hingegen von einer angemessen durchgeführten Planungsphase aus, die jedoch nicht dokumentiert wurde, hätten die Gutachter auch an dieser Stelle gegen das wissenschaftliche Gebot der Transparenz verstoßen.

5.2. Datenerhebung, -auswertung und Ergebnisdarstellung

Die Qualität der *Datenerhebung, -auswertung und Ergebnisdarstellung* wurde in allen analysierten Gutachten als mangelhaft befunden. Da vorliegend explizite Hinweise auf eine Planung der jeweiligen Untersuchung fehlten, überrascht es wenig, dass der

in den Gutachten beschriebene Prozess der Datenbeschaffung der Vorgehensweise eines „materialschaffenden Experiments“ (Holzkamp, 1964) ähnelt: Die Daten aus den mündlichen Befragungen, Beobachtungen und testpsychologischen Verfahren werden erst einmal – zur späteren Entscheidungsfindung – gesammelt. Dieses Vorgehen ist ineffizient und auch ethisch fragwürdig. Darüber hinaus entspricht die Dokumentation der Datenbeschaffung nicht wissenschaftlichen Ansprüchen. Diese wurden bereits 1964 von Holzkamp ausgeführt: „Erfahrungswissenschaftliches Vorgehen will objektiv sein, in dem Sinne, dass in dieser Weise beschafftes Wissen möglichst gut kommunizierbar, d. h. von möglichst wenig Mehrdeutigkeiten belastet sein soll. Das macht erforderlich, die Umstände der Wissensbeschaffung möglichst präzise und konkret festzustellen und mitzuteilen. ... Jedermann muss gleichsam in die Lage versetzt werden, das irgendwo gewonnene Wissen selbst nachzukonstruieren, in der Wirklichkeit zu reproduzieren“ (s. hierzu Holzkamp, 1964). Vorliegend lässt keine Aufzeichnung der Untersuchungsdurchführung eine Reproduktion der Diagnosen bei gegebenen Daten zu. Zudem ist aufgrund der fehlenden Angaben zu standardisierten Instruktionen und Verfahrenshinweisen zur Durchführung der Datenerhebung und -auswertung fraglich, wieweit in den psychiatrisch-psychologischen Untersuchungen eine einheitliche Durchführung, Klassifikation und Verrechnung der erhobenen Daten gewährleistet waren. Falls tatsächlich keine standardisierten Vorgaben für die Befragungen und Beobachtungen existierten, hätten Tür und Tor weit offen gestanden für Prozessfehler (z. B. Rammstedt et al., 2014), da von einer erheblichen Abhängigkeit der erhobenen Daten von Ermessensentscheidungen der gutachtenden Personen auszugehen wäre. Gab es hingegen Standards, die aber nicht mitgeteilt wurden, wären für die Gutachten auch an dieser Stelle mangelnde Sorgfalt und Transparenz festzustellen.

Aus den mündlichen Befragungen und Beobachtungen resultierten intransparente Ergebnisse. Vergleichbare Qualitätsmängel berichteten Nowara (1995) und Wulf (2005). Die Ergebnisdarstellung der psychologischen Testdiagnostik offenbarte in den vorliegenden Gutachten defizitäre Methodenkenntnisse der Gutachter: Der Messfehler blieb in der Ergebnisdarstellung und -interpretation unberücksichtigt, Vergleichs- und Referenzgrößen wurden nicht mitgeteilt, Entscheidungsgrundlagen zur Beurteilung der Fehlerrisiken nicht genannt. Diese Defizite, die Nedopil bereits 2005 feststellte, sind als gravierend zu werten, da die testpsychologischen Befunde bei einem Teil der Gutachten für die finalen Prognosen herangezogen wurden, die reale

Entscheidungen implizierten, wie bspw. klassifikatorische Zuordnungen, die ihrerseits wiederum Konsequenzen nach sich zogen, z. B. für selektive Indikationsentscheidungen bezüglich des Arbeitsplatzes oder einer therapeutischen Maßnahme.

5.3. Datenintegration und Empfehlung von Maßnahmen

Vorliegend war die Qualität der *Interpretation und Integration der Daten* sowie die der *Empfehlung von Maßnahmen* mangelhaft (31 % der Gutachten) bis ausreichend (69 % der Gutachten). Die Urteilsbildung entbehrte einer transparenten, empirisch informierten Wissensbasis. Sie erfolgte klinisch-intuitiv, ein Ergebnis, das mit den Befunden von Kunzl und Pfäfflin (2011) sowie Haubner-Maclean und Eher (2014) übereinstimmt. Für diese, ausschließlich auf der professionellen Expertise der Gutachtenden beruhenden Urteilsbildung wurden robust Mängel aufgezeigt, darunter eine Anfälligkeit für universelle Verzerrungs- und Fehlertendenzen der menschlichen Beurteilung (Oberlader & Schmidt, 2024; Rettenberger & Eher, 2016), eine mangelnde Reliabilität der psychiatrischen und klinisch-psychologischen Diagnosen (Spitzer & Fleiss 1974; Megargee 1976), Klassifikationsfehler zugunsten von falsch-positiven Einschätzungen (Monahan, 1981), eine inkonsistente Verwendung von diagnostischen Entscheidungsstrategien (Kuncel et al., 2013), eine unzureichende Prädiktionsleistung und ein intransparenter Urteilsprozess (Grove et al., 2000; Grove & Meehl, 1996; s. hierzu auch Holtmann et al., 2020; Rettenberger, 2018)) Aufgrund ihrer erheblichen Fehleranfälligkeit sollte diese Prognosepraxis, „eigentlich für eine wissenschaftlich fundierte Expertise, die einem akademischen Anspruch folgt, keine Rolle mehr spielen“ (Rettenberger, 2023, S. 2). Die empfohlenen Maßnahmen fußten fast ausschließlich auf dem (Nicht-)Vorliegen von Störungsdiagnosen. Diese Basis ist schmal, von einer begrenzten prognostischen Relevanz (Biedermann et al., 2023) und bleibt hinter den erfahrungswissenschaftlichen Empfehlungen für Prognosegutachten zurück. Da auch im vorliegenden Kontext vom Garbage in-Garbage out-Prinzip auszugehen ist, können im Ergebnis des beschriebenen gutachterlichen Vorgehens keine wissenschaftlich fundierten Prädiktionen resultieren. Ob sie deshalb *falsch* sind, ist eine andere, inhaltliche und empirisch zu prüfende Fragestellung.

5.4. Zusammenfassende Würdigung, Implikationen und Limitationen

Zusammenfassend legen die dargestellten Befunde nahe, dass die empfohlenen Mindeststandards für Prognosegutachten in den vorliegend analysierten

Begutachtungsprozessen nicht in der gebotenen Weise umgesetzt wurden. Die Ergebnisse verdeutlichen insbesondere die gravierende Fehlerkaskade, die eine unangemessene Konzeptualisierung und eine verfehlte Planung der psychiatrisch-psychologischen Untersuchung nach sich ziehen können. Versäumt wurde eine transparente Begutachtungspraxis mit empirisch begründeten Empfehlungen (APA 2006), die auf einer wissenschaftlich fundierten Konzeptualisierung und Planung der psychiatrisch-psychologischen Untersuchung gründet, individuelle Risiko- und Schutzfaktoren berücksichtigt (Verschuere et al., 2023), Regeln für die Erstellung der Diagnosen bzw. Prognosen expliziert, Fehlerrisiken beurteilt, das Kostenverhältnis der möglichen Handlungsoptionen abwägt (Sterner et al., 2024) und auch die Gelingensbedingungen der empfohlenen Maßnahmen aufzeigt (Philipp & Rettenberger, 2025). Angesichts der weitreichenden Konsequenzen, welche die Prognosestellungen für die begutachteten Kleriker, aber auch für die zu schützenden Minderjährigen haben können, ist dieser Befund inakzeptabel.

Gleichzeitig bedeutet dieser Befund, dass die vorliegend analysierten Gutachten keine verlässlichen Datenquellen sind für eine inhaltsanalytische Rekonstruktion des Missbrauchsgeschehens aus Beschuldigtenperspektive.

Darüber hinaus verdeutlicht das vorliegende Ergebnis die Notwendigkeit einer systematischen Qualitätskontrolle der Begutachtung von pädosexuellen Beschuldigten und Tätern im Bistum Trier. In diesem Zusammenhang ist auf einen qualitätssichernden Handlungsbedarf auf Seiten des Auftraggebers hinzuweisen. Er sollte diejenigen Sachverständigen auswählen, die hinreichend qualifiziert sind und ihre fachlichen Kenntnisse in der Begutachtungspraxis angemessen, im Sinne eines wissenschaftlich fundierten Vorgehens, anwenden. Dies setzt voraus, dass der Auftraggeber über die notwendigen Kompetenzen verfügt, die fachliche Expertise der Gutachter angemessen zu beurteilen.

Abschließend ist ausdrücklich darauf hinzuweisen, dass die vorliegenden Ergebnisse keine Aussagen darüber zulassen, ob die identifizierten Qualitätsmängel auf die Begutachtungspraxis anderer Bistümer verallgemeinerbar sind, ob sie auf die Begutachtung pädosexueller Täter generell übertragen werden dürfen und, ob sie weiterführend auch auf andere Begutachtungsfelder zutreffen.

6. Literatur

- American Psychological Association (2013). Specialty guidelines for forensic psychology. *The American psychologist*, 68(1), 7–19.
<https://doi.org/10.1037/a0029889>
- American Psychological Association, Presidential Task Force on Evidence-Based Practice. (2006). Evidence-based practice in psychology. *American Psychologist*, 61(4), 271–285. <https://doi.org/10.1037/0003-066X.61.4.271>
- American Academy of Psychiatry and the Law. (2005). *Guidelines and practice resources*. American Academy of Psychiatry and the Law.
<https://www.aapl.org/guidelines-and-practice-resources>
- Beckmann, D., & Richter, H.-E. (1991). *Der Gießen-Test (GT): Manual*. Huber.
- Berufsethische Richtlinien des Berufsverbandes Deutscher Psychologinnen und Psychologen e. V. und der Deutschen Gesellschaft für Psychologie e. V. (2022). *Berufsethische Richtlinien*. Berufsverband Deutscher Psychologinnen und Psychologen e. V. & Deutsche Gesellschaft für Psychologie e. V.
- Biedermann, L., Eher, R., Rettenberger, M., Gaunersdorfer, K., & Turner, D. (2023). Are mental disorders associated with recidivism in men convicted of sexual offenses? *Acta Psychiatrica Scandinavica*. Advance online publication.
<https://doi.org/10.1111/acps.13547>
- Boetticher, A., Kröber, H.-L., Müller-Isberner, R., Böhm, K. M., Müller-Metz, R., & Wolf, T. (2007). Mindestanforderungen für Prognosegutachten. *Forensische Psychiatrie, Psychologie, Kriminologie*, 1, 90–100.
- Boetticher, A., Koller, M., Böhm, K. M., Brettel, H., Dölling, D., Höffler, K., Müller-Metz, R., Pfister, W., Schneider, U., Schöch, H., & Wolf, T. (2019). Empfehlungen für Prognosegutachten: Rechtliche Rahmenbedingungen für Prognose im Strafverfahren. *Forensische Psychiatrie, Psychologie, Kriminologie*, 13(4), 305–333.
- Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6. Aufl.). Anderson Publishing.
- Cronbach, L. J., & Gleser, G. C. (1965). *Psychological tests and personnel decisions*. University of Illinois Press.

- Dahle, K. P. (2005). Psychologische Begutachtung zur Kriminalprognose. In H.-L. Kröber & M. Steller (Hrsg.), *Psychologische Begutachtung im Strafverfahren*.
- Dahle, H. P., & Lehmann, R. J. B. (2018). Zum prognostischen Mehrwert einer integrativen nomothetisch idiografischen kriminalpsychologischen Prognosebeurteilung – Eine empirische Untersuchung an männlichen Gewalt- und Sexualstraftätern. *Forensische Psychiatrie, Psychologie, Kriminologie*, 12(1), 37–50.
- Diagnostik- und Testkuratorium der Föderation Deutscher Psychologenvereinigungen. (2017). *Qualitätsstandards für psychologische Gutachten*.
https://www.dgps.de/fileadmin/documents/Empfehlungen/GA_Standards_DTK_1Sep_2017_Final.pdf
- Dreßing, H., & Foerster, K. (2021). Traumafolgestörungen in ICD-10, ICD-11 und DSM-5: Diagnosekriterien und ihre Bedeutung für die gutachtliche Praxis. *Forensische Psychiatrie, Psychologie, Kriminologie*, 15(1), 47–53.
<https://doi.org/10.1007/s11757020-00645-6>
- Egil-Alge, M. (2017). Risikoprognose- Schicksal und Chance? In U. Kobbé (Hrsg.), *Forensische Prognosen: Ein transdisziplinäres Praxismanual. Standards, Leitfäden, Kritik* (S. 339 - 348). Pabst.
- Engel, R. R. (2000). *Minnesota Multiphasic Personality Inventory – 2 (MMPI-2): Handanweisung*. Hogrefe.
- Fegert, J. M., Schnoor, K., König, C., & Schläfke, D. (2006). *Psychiatrische Begutachtung in Sexualstraßverfahren: Eine empirische Untersuchung von Gutachten zur Schuldfähigkeit bei jugendlichen, heranwachsenden und erwachsenen Beschuldigten in Mecklenburg- Vorpommern*. Centaurus.
- Foerster, K. (2003). Von der Verantwortung des psychiatrischen Sachverständigen. In K. Amelung (Hrsg.), *Strafrecht – Biorecht – Rechtsphilosophie: Festschrift für Hans-Ludwig Schreiber zum 70. Geburtstag am 10. Mai 2003* (S. 81–88). C. F. Müller.
- von Franqué, F. (2013). Strukturierte, professionelle Risikobeurteilungen. In M. Rettenberger und F. von Franqué (Hrsg.), *Handbuch kriminalprognostischer Verfahren*. (S. 357-380).
- Galli, T. (2011). *Lockerungsbegutachtungen im Strafvollzug*. Springer Science & Business Media.

- Graumann, C. F. (1966). *Bewußtsein und Bewußtheit Probleme und Befunde der psychologischen Bewußtseinsforschung*. Verlag für Psychologie.
- Gretenkord, L. (2017). Ist die (vorzeitige) Entlassung verantwortbar? Sieben Schritte zu einer wissenschaftlichen, juristischen und ethisch fundierten Entscheidung. In U. Kobbé (Hrsg.), *Forensische Prognosen: Ein transdisziplinäres Praxismanual. Standards, Leitfäden, Kritik* (S. 87 - 98). Pabst.
- Grewe, W., & Wentura, D. (1997). *Wissenschaftliche Beobachtung: Eine Einführung* (2. Aufl.). Psychologie Verlags Union.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30. <http://dx.doi.org/10.1037/10403590.12.1.19>
- Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, 2(2), 293–323. <https://doi.org/10.1037/1076-8971.2.2.293>
- Hock, M., Peters, J., Renner, K.-H. & Krohne, H. W. (2023). *Psychologische Diagnostik: Grundlagen und Anwendungsfelder* (3. Aufl.). Stuttgart: Kohlhammer.
- Haubner-Maclean, T., & Eher, R. (2014). Nicht mehr gefährlich und doch rückfällig? Die ungenügende Abbildung gefährlichkeitsrelevanter Merkmale bei rückfälligen ehemals untergebrachten Sexualstraftätern. *Recht & Psychiatrie*, 32(2), 69–79.
- Holtmann, J., Koch, T., Bohn, J., & Eid, M. (2020). Multi-method assesement of time-stable and time-variable interindividual differences. *European Journal of Psychological Assessment*, 36(6), 1024–1043.
- Holzkamp, K. (1964). *Theorie und Experiment in der Psychologie: Eine grundlagenkritische Untersuchung*. De Gruyter.
- Kaminski, G. (1970). *Verhaltenstheorie und Verhaltensmodifikation*. Klett.
- Kane, M. T. (2013). Validating the interpretations and the uses of test scores. *Journal of Educational Measurement*, 50, 1–73.

- Kanning, U. P. (2022). Systematische Fehler der Personenbeurteilung. In R. Dohrenbusch (Hrsg.), *Psychologische Begutachtung*. Springer.
- Kindler, H., & Fegert, J. M. (2023). Auswertung und Qualitätsprüfung von Sachverständigengutachten. In J. M. Fegert, T. Meysen, H. Kindler, K. Chauviré-Geib, U. Hoffmann, & E. Schumann (Hrsg.), *Gute Kinderschutzverfahren* (S. 583–606). Springer.
- Kochinka, A. (2010). Beobachtung. In G. Mey & K. Mruck (Hrsg.), *Handbuch Qualitative Forschung in der Psychologie* (S. 449–461). Springer VS.
- Krippendorff, K. (1970). Estimating the reliability, systematic error, and random error of interval data. *Educational and Psychological Measurement*, 30(1), 60–70. <https://doi.org/10.1177/001316447003000105>
- Krippendorff, K. (2019). *Content analysis*. SAGE Publications.
- Krüger, H.-L., Brettel, H., Rettenberger, M., & Stübner, S. (2019). Empfehlungen für Prognosegutachten: Erfahrungswissenschaftliche Empfehlungen für kriminalprognostische Gutachten. *Forensische Psychiatrie, Psychologie, Kriminologie*, 13, 334–342. <https://doi.org/10.1007/s11757-019-00558-z>
- Kunzl, F., & Pfäfflin, F. (2011). Qualitätsanalyse österreichischer Gutachten zur Zurechnungsfähigkeit und Gefährlichkeitsprognose. *Recht & Psychiatrie*, 29(3), 152–159.
- Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, 98(6), 1060–1072.
- Kurtz, C. (2017). *Meta-analytische Studie zur Prognose von kriminellen Rückfällen auf der Basis empirischer Daten*. Pabst Science Publishers.
- Kury, H., & Adams, B. (2010). Prognosegutachten im Strafvollzug. *Forum Strafvollzug*, 59(2), 81–87.
- Megargee, E. J. (1976). The prediction of dangerous behavior. *Criminal Justice and Behavior*, 3(1), 3–21. <https://doi.org/10.1177/009385487600300101>
- Meehl, P. E. (2013). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. University of Minnesota Press. (Erstveröffentlichung 1954).

- Monahan, J. (1981). *The clinical prediction of violent behavior*. U.S. Government Printing Office.
- Murrie, D. C., & Warren, J. I. (2005). Clinician variation in rates of legal sanity opinions: Implications for self-monitoring. *Professional Psychology: Research and Practice*, 36(5), 519-524. <https://doi.org/10.1037/0735-7028.36.5.519>
- Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie: Ein Handbuch für die Praxis*. Pabst Science Publishers.
- Nedopil, N. (2022). Forensisch-psychiatrische Risikoeinschätzung. In F. Häßler, N. Nedopil, & M. Dudeck (Hrsg.), *Praxishandbuch Forensische Psychiatrie: Grundlagen, Begutachtung, Interventionen im Erwachsenen-, Jugendlichen- und Kindesalter* (3. Aufl., S. 609–640). Medizinisch-Wissenschaftliche Verlagsgesellschaft.
- Nowara, S. (1995). Externe Prognosegutachten im Maßregelvollzug. *Recht & Psychiatrie*, 13, 67–72.
- Philipp, F., & Rettenberger, M. (2025). Mindestanforderungen/Empfehlungen für Prognosegutachten. In R. Dohrenbusch (Hrsg.), *Psychologische Begutachtung* (S. 1–10). Springer.
- Prüter-Schwarte, C. (2022). Aufbau des schriftlichen Gutachtens. In F. Häßler, N. Nedopil und M. Dudeck (Hrsg.). *Praxishandbuch Forensische Psychiatrie. Grundlagen, Begutachtung, Interventionen im Erwachsenen-, Jugendlichen- und Kindesalter*, (3. Aufl.), (S. 73-75). Medizinische Wissenschaftliche Verlagsgesellschaft.
- Prüter-Schwarte, C., Lutz, M., & Kunert, H. J. (2019). Die Qualität von Schuldfähigkeitsgutachten im Maßregelvollzug untergebrachter Patienten – Ergebnisse einer empirischen Studie. *Der Medizinische Sachverständige*, 115(5), 203–207.
- Rammstedt, B., Beierlein, C., Brähler, E., et al. (2014). Qualitätsstandards zur Entwicklung, Anwendung und Bewertung von Messinstrumenten in der sozialwissenschaftlichen Umfrageforschung. *Journal of Contextual Economics – Schmollers Jahrbuch*, 134(4), 517-546. <https://doi.org/10.3790/schm.134.4.517>

- Rettenberger, M. (2017). Die kriminalprognostische Einschätzung und Begutachtung bei Kindesmissbrauch. In T. Stompe & H. Schanda (Hrsg.), *Sexueller Kindesmissbrauch und Pädophilie: Grundlagen, Begutachtung, Prävention und Intervention – Täter und Opfer* (S. 311–323). Medizinisch-Wissenschaftliche Verlagsgesellschaft.
- Rettenberger, M. (2018). Intuitive, klinisch-idiographische und statistische Kriminalprognosen im Vergleich - die Überlegenheit wissenschaftlich strukturierten Vorgehens. *Forensische Psychiatrie, Psychologie, Kriminologie*, 12(1), 28–36. <https://doi.org/10.1007/s11757-017-0463-y>
- Rettenberger, M. (2023). Methoden der Kriminalprognosebegutachtung. In: Dohrenbusch, R. (Hrsg.). *Psychologische Begutachtung*. Springer. https://doi.org/10.1007/978-3-662-64801-8_125-1
- Rettenberger, M., & Brettel, H. (2020). Qualitätssicherung in der Kriminalprognostik: Die neuen Mindestanforderungen („Empfehlungen“) für Prognosegutachten. *Praxis der Psychologie und des Rechts*, 30(1), 29–42.
- Rettenberger, M., & von Franquè, F. (Hrsg.). (2013). *Handbuch kriminalprognostischer Verfahren*. Hogrefe.
- Salewski, C., & Stürmer, S. (2014). *Qualitätsmerkmale in der familienrechtspsychologischen Begutachtung. Untersuchungsbericht I*. Fern Universität Hagen. https://www.fernuni-hagen.de/psychologie/docs/untersuchungsbericht1_fprgutachten014.pdf
- Segmiller, F., & Dudeck, M. (2019). *Psychiatrische Beispielgutachten: Ein Praxisbuch für Einsteiger und Fortgeschrittene*. Kohlhammer.
- Spitzer, R. L., & Fleiss, J. L. (1974). A re-analysis of the reliability of psychiatric diagnoses. *The British Journal of Psychiatry*, 125(3), 341–347. <https://doi.org/10.1192/bjp.125.3.341>
- Steinböck, H. (2017). Ein Jahr Bayerisches Maßregelvollzugsgesetz -eine erste Bilanz aus der klinischen Praxis. *Recht & Psychiatrie*, 35(3).
- Sterner, P., Friemelt, B., Goretzko, D., Kraus, E., Bühner, M., & Pargent, F. (2024). Das Konfidenz- / Signifikanzniveau impliziert ein bestimmtes Kostenverhältnis zwischen Fehler 1. Art und Fehler 2. Art: Für ein stärkeres Einbeziehen

- der Entscheidungstheorie in die psychologische Einzelfalldiagnostik. *Diagnostica*, 70(3), 126–138. <https://doi.org/10.1026/0012-1924/a000329>
- Stiels-Glenn, M. (2017). Prognose mit Straftätern statt über sie. Methodische und ethische Fragen. In U. Kobbé (Hrsg.), *Forensische Prognosen: Ein transdisziplinäres Praxismanual. Standards, Leitfäden, Kritik* (S.349 - 356). Pabst.
- Tack, W. H. (1976). Diagnostik als Entscheidungshilfe. In K. Pawlik (Hrsg.), *Diagnose der Diagnostik*. Klett.
- Verschuere, B., Lin, C. C., Huismann, S., et al. (2023). The use-the-best heuristic facilitates deception detection. *Nature Human Behaviour*, 7(5), 718–728.
- Wertz, M., & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale. *Forensische Psychiatrie, Psychologie und Psychotherapie*, 28(3), 241–261. https://doi.org/10.1486/RP-2020-04_193
- Wertz, M., Kury, H., & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis: Eine empirische Validierung unter Berücksichtigung der Rückfallquoten. *Forensische Psychiatrie, Psychologie, Kriminologie*, 12(1), 51 - 60. <https://doi.org/10.1007/s11757-017-0458-8>
- Westhoff, K., & Kluck, M. L. (2013). *Psychologische Gutachten schreiben und beurteilen*. Springer.
- Wulf, R. (2005). Gute kriminologische Prognose: Rückfall, Flucht, Suizid. *Monatschrift für Kriminologie und Strafrechtsreform*, 88(4), 290–304.
- Yoon, D., Turner, D., Klein, V., Rettenberger, M., Eher, R., & Briken, P. (2018). Factors predicting desistance from reoffending: A validation study of the SAPROF in sexual offenders. *International Journal of Offender Therapy and Comparative Criminology*, 62(3), 697–716.