Quality of Criminal Risk Assessment –

Empirical Studies about the Importance of Minimum Requirements and Different Methodological Approaches for Criminal Risk Assessment

Inauguraldissertation

zur Erlangung des Akademischen Grades

eines Dr. phil.,

vorgelegt dem

Fachbereich 02 - Sozialwissenschaften, Medien und Sport

der Johannes Gutenberg-Universität Mainz

von

Maximilian Wertz

aus Menden (Sauerland)

Mainz

2022

Referent: Prof. Dr. Martin Rettenberger

1. Koreferent: Prof. Dr. Dr. Hauke Brettel

2. Koreferentin: Prof. Dr. Dudeck

Tag des Prüfungskolloquiums:

Erklärung

Kapitel 2 beruht auf folgendem Manuskript:

Wertz, M., Kury, H. & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [The application of methodological minimum requirements for risk assessment reports in clinical practice - An empirical validation using officially registered reoffenses]. *Forensische Psychiatrie, Psychologie, Kriminologie, 12*(1), 51 – 60.

Kapitel 3 beruht auf folgendem Manuskript:

Wertz, M., Schiltz, K., Imhoff, R.& Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosebegutachtung [The influence of the judicial order on the quality of the work of expert witnesses in the context of risk assessment]. *Recht & Psychiatrie, 38*(4), 193 – 200.

Kapitel 4 beruht auf folgendem Manuskript:

Wertz*, M., Hausam*, J., Konrad, N., Schiltz, K., Imhoff, R.& Rettenberger, M. (2021). Qualität von Schuldfähigkeitsgutachten – Mindestanforderungen, unterbringungsrelevante Gefährlichkeitsprognose und Berücksichtigung im richterlichen Urteil [Quality of criminal responsibility reports – Minimum requirements, risk assessment, and consideration in court decisions]. *Recht & Psychiatrie, 39*(4), 202 – 211. * geteilte Erstautorenschaft

Kapitel 5 beruht auf folgendem Manuskript:

Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R. & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. *Psychology, Crime & Law*. *Advance online publication.* https://doi.org/10.1080/1068316X.2022.2063286

Kapitel 6 beruht auf folgendem Manuskript:

Wertz, M., & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale [The use of standardized risk assessment instruments in the practice of risk assessment: Empirical findings on frequency and risk communication as a function of assessment- and subject-related characteristics.]. *Forensische Psychiatrie und Psychotherapie, 28*(3), 241-261.

Kapitel 7 beruht auf folgendem Manuskript:

Wertz, M., Schobel, S., Schiltz, K. & Rettenberger, M. (in press). A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. *Psychological Assessment.*

Kapitel 1 sowie Kapitel 8 enthalten Teile der oben genannten Manuskripte.

Abstract

Risk assessment reports about individuals convicted of sexual and/or violent offenses are an issue of increasing public and socio-political concern, and their (methodological) quality is regularly and intensively discussed also in research literature. Since the publication of methodological minimum requirements for risk assessment reports of an interdisciplinary working group in 2006 in Germany and updated recommendations in 2019, there is little empirical evidence on whether and how these standards are put into clinical and judicial practice. In an iterative research project of six independent empirical studies, a systematic retrospective analysis of more than 1.000 risk and criminal responsibility assessment reports from the penitentiary in Freiburg, the Forensic Psychiatry of the Charité in Berlin, and the Department of Forensic Psychiatry of the University Hospital Munich from 1999 to 2015 was conducted. Based on this, judicial verdicts, and officially registered re-offenses according to the Federal Central Register (June 2016) were examined and related to the degree of application of the above-mentioned minimum requirements, different methodological approaches to criminal risk assessment, and the use of psychometric tests and structured criminal risk assessment instruments (in both risk and criminal responsibility reports). Summarising, the results showed an increasing implementation of minimum requirements and standardization of clinical and judicial criminal risk assessment practice, providing further support for the use of structured and standardized risk assessment procedures compared to unstructured approaches. On the one hand, the results indicate a (partial) positive effect, on the other hand, more efforts are needed regarding further quality assurance of both criminal risk and responsibility assessments. The results are discussed from different perspectives, and implications for practice are given.

Keywords: criminal risk assessment, criminal responsibility assessment, methodological minimum requirements, recidivism, expert reports

Zusammenfassung

Kriminalprognostische Gutachten über Gewalt- und Sexualstraftäter stehen zunehmend im öffentlichen und gesellschaftspolitischen Fokus und sehen sich auch in der Forschungsliteratur einer intensiveren Diskussion hinsichtlich ihrer Qualität ausgesetzt. Seit der Veröffentlichung von Mindestanforderungen für Prognosegutachten einer interdisziplinären Arbeitsgruppe im Jahr 2006 und deren Fortschreibung als Empfehlungen im Jahr 2019 liegen bislang nur wenige empirische Belege darüber vor, ob und in welcher Form diese auch in der gutachterlichen und richterlichen Praxis umgesetzt werden. Im Rahmen eines iterativen Forschungsprojekts sechs empirischer Einzelstudien wurden über 1.000 Prognose- und Schuldfähigkeitsgutachten aus der Justizvollzugsanstalt Freiburg, der Charité – Universitätsmedizin Berlin und der Abteilung für Forensischen Psychiatrie der Klinik und Poliklinik für Psychiatrie und Psychotherapie der Ludwig-Maximilians-Universität München von 1999-2015 systematisch analysiert. Darauf aufbauend wurden Auskünfte über die gerichtlichen Verfahrensausgänge sowie die offiziellen Rückfalldaten der Probanden laut Bundeszentralregisterauszug (Stand Juni 2016) ausgewertet und in Kontext mit der Einhaltung der formulierten Mindestanforderungen, unterschiedlicher Ansätze der Prognosemethodik und dem Einsatz psychologischer Testverfahren bzw. Risikoinstrumente (in Prognose- und Schuldfähigkeitsgutachten) gesetzt. Es zeigte sich zusammenfassend eine zunehmende Umsetzung der Mindestanforderungen sowie Standardisierung in der gutachterlichen und richterlichen (kriminalprognostischen) Praxis, wobei eine signifikant höhere Vorhersageleistung durch standardisierte, strukturierte Prognoseansätze gegenüber unstrukturierter Urteilsbildung empirisch belegt werden konnte. Die Ergebnisse sprechen einerseits für einen (Teil-)Erfolg, andererseits verdeutlichen sie weiteren Handlungsbedarf im Hinblick auf die Qualitätssicherung bei der Erstellung von Schuldfähigkeits- und Prognosegutachten. Die Ergebnisse werden aus verschiedenen Perspektiven diskutiert und es werden Implikationen für die Praxis gegeben.

Table of Contents

1. General Introduction[1]

Criminal risk assessments of individuals already convicted or charged of offenses are highly relevant and one of the most important tasks of legal decision-makers in German criminal law. Prognostic decisions entail far-reaching freedom-related consequences for individuals (Rettenberger, 2018; 2019), as they significantly influence the (duration of) imprisonment or accommodation in a psychiatric forensic hospital (Müller & Nedopil, 2017; Prüter-Schwarte et al., 2019; Verrel 1995; 2015).

In the case of legal decisions of considerable consequence, experts from psychoscientific professions are regularly consulted (Boetticher et al., 2019) in order to ensure a scientifically sound basis for the legal decisions (Dahle, 2010). Risk assessments, in addition to assisting with potential granting of privileges or release measures for individuals convicted of offenses, also serve as a basis for assessing the prerequisites of a custodial measure in the context of criminal responsibility assessments. Criminal risk assessments thus represent an essential component of the areas of responsibility of persons working in penitentiaries or forensic psychiatry's or of external expert witnesses (Basdekis-Jozsa, et al., 2013; Dahle, 2005; Gretenkord, 2013). Prognostic decisions are also an important component

[1] Parts of the Introduction are published in Wertz, M., Kury, H. & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [The application of methodological minimum requirements for risk assessment reports in clinical practice - An empirical validation using officially registered reoffenses]. Forensische Psychiatrie, Psychologie, Kriminologie, 12(1), 51 – 60; Wertz, M., Schiltz, K., Imhoff, R.& Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosebegutachtung [The influence of the judicial order on the quality of the work of expert witnesses in the context of risk assessment]. Recht & Psychiatrie, 38(4), 193 – 200; Wertz, M., Hausam, J., Konrad, N., Schiltz, K., Imhoff, R.& Rettenberger, M. (2021). Qualität von Schuldfähigkeitsgutachten – Mindestanforderungen, unterbringungsrelevante Gefährlichkeitsprognose und Berücksichtigung im richterlichen Urteil [Quality of criminal responsibility reports – Minimum requirements, risk assessment, and consideration in court decisions]. Recht & Psychiatrie, 39(4), 202 – 211; Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R. & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. Psychology, Crime & Law. Advance online publication. https://doi.org/10.1080/1068316X.2022.2063286; Wertz, M., & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale [The use of standardized risk assessment instruments in the practice of risk assessment: Empirical findings on frequency and risk communication as a function of assessment- and subject-related characteristics.]. Forensische Psychiatrie und Psychotherapie, 28(3), 241-261, and in Wertz, M., Schobel, S., Schiltz, K. & Rettenberger, M. (in press). A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. Psychological Assessment.

of planning and implementing the most effective risk management measures possible (Rettenberger, 2019), the naming and discussion of which are explicitly called for in current recommendations for risk assessment reports (Boetticher et al., 2019). In addition to general remarks on risk assessment and possible treatment options, the formulated requirements also address whether and how a persisting criminal risk can be controlled or reduced by appropriate risk management measures.

To date, research literature of the last few decades points to a relatively heterogeneous quality of expert witness reports (Nowara, 1995a, 1995b; Suhling, 2003; Dahle et al., 2009, 2012; Haarig et al., 2012; Kunzl & Pfäfflin, 2011). Since the publication of methodological minimum requirements for risk assessment reports in 2006 in Germany and updated recommendations in 2019, there is little empirical evidence on whether and how these standards are put into clinical and judicial practice. It also remains unclear, how frequently psychometric tests and actuarial and clinical-structured prognostic instruments are used in real criminal assessment practice in Germany. Given the frequently reported superiority of structured risk assessment approaches compared to unstructured and intuitively made judgments, various aspects of standardization in this field of assessment research and practice are still discussed controversially.

Therefore, the aim of this thesis was an extensive examination of the quality of criminal risk assessment reports about individuals convicted or charged of sexual and/or violent offenses in German practice, focusing especially on the importance of published minimum requirements and different methodological approaches for criminal risk assessment. For this purpose, a systematic retrospective analysis of more than 1.000 risk and criminal responsibility assessment reports from the penitentiary in Freiburg, the Forensic Psychiatry of the Charité in Berlin, and the Department of Forensic Psychiatry of the University Hospital Munich from 1999 to 2015 was conducted. Based on this, judicial verdicts, and officially registered re-offenses according to the Federal Central Register (June 2016) were examined

and related to the degree of application of the above-mentioned minimum requirements by witness experts (see Chapter 2) and judicial orders (see Chapter 3), the use of psychometric tests (see Chapter 5) and structured criminal risk assessment instruments (see Chapter 6), as well as different methodological approaches to criminal risk assessment (see Chapter 7). As an excursus, the implementation of criminal risk assessments in and quality of criminal responsibility reports were additionally examined (see Chapter 4).

Before the six independent empirical studies are described in more detail, an overview will be given of the research literature regarding quality, the different methodological approaches, and efforts regarding further quality assurance and standardization of criminal risk and responsibility assessments. In addition, the legal framework of expert witnesses in the German Penal Law will be initially described for a more comprehensive background understanding.

1.1 Legal Background of Expert Witnesses in the German Penal Law

In Germany, the majority of criminal assessment reports focus on the responsibility and recidivism risk of individuals charged or convicted of violent and/or sexual offenses (Dunn et al., 2014). Criminal responsibility assessment reports are required to assess whether there is a clinical diagnosis which abolishes or diminishes the insight of the alleged person regarding the wrongfulness of the criminal act or the capability to act through a lack of insight, and if the offender poses a risk of future offences (due to his or her assumed mental disorder). The German legal system has three categories of responsibility in mentally disordered offenders: responsible, diminished responsible, and not responsible (articles 20 and 21 of the German Penal Code). The forensic assessment requires a two-step evaluation: first, the alleged person must have a legally defined disorder at the time of the offense. These respective legal terms are: Severe Mental Disease, Severe Disturbances of Consciousness, Mental Retardation, and Other Severe Disturbances of the Mind (translations as close as possible to the legal code terminology; for an empirical evaluation of the stigmatizing nature

of these terms see Rösch et al., 2021). Mental Diseases in the sense of the law are, for example, organic or genetic disorders, diseases of the brain, severe acute intoxications, and major mental disorders. The term Severe Disturbances of Consciousness means short but exceptionally severe mental disturbances in individuals, who are otherwise mentally sane (i.e., severely affective exceptional situations). The term refers to the evaluation of criminal responsibility regarding so-called affect crimes due to consciousness disturbances. Mental retardation is understood as a severe deficit of intelligence, which restricts substantially the ability to understand and act due to legal standards. All other mental disorders, like personality disorders, sexual deviations, and other chronic or longer lasting reactive disorders are classified with the term Other Severe Disturbances of the Mind.

If the defendant suffered from one of these disorders, the expert witness must assess the functional consequences of the disorder at the moment of the criminal act. The alleged person, who was unable to understand the unlawfulness of the criminal act or lacked the ability to control his actions because of one of the four disorders mentioned in article 20, is usually considered not responsible. When his or her ability to control his or her actions was severely diminished, the responsibility is regarded as diminished (article 21 of the penal code). Finally, psychologists and psychiatrists will also be asked about the dangerousness of a defendant. Defendants who are not responsible and who are at the same time considered to be dangerous are sentenced to mandatory treatment in a forensic mental hospital. Those who are diminished responsible are sentenced to both hospital treatment and prison, and the time spent in prison will be deducted from the period of the prison sentence. Individuals suffering from addiction disorders are sentenced by article 64 of the German penal code and individuals suffering from any other diagnoses by article 63 of the German penal code.

Furthermore, Germany also has provisions for preventive detention of persons regarded as dangerous serial offenders (article 66 of the German penal code). Risk assessments from mental health professionals will be required not only at the time of

sentencing but several times thereafter, because detaining and releasing mentally ill or chronically dangerous offenders from hospitals or prisons depend strongly on the results of their risk assessment reports. The mental health professionals do not make decisions about detaining or releasing a person but give their advice to the courts, which then have to base their decisions upon their legal understanding and the empirical evidence of the individual case (Dunn et al., 2014).

Not just regularly but in the majority of cases in Germany, psychiatric hospital institutions and penitentiaries are asked to conduct such assessment reports. University hospital institutions provide external forensic assessment reports for a diverse number of different courts or public prosecutors as well as education and training of forensic students, psychologists, psychiatrists, and other clinicians working in the forensic field. Penitentiaries are run by federal states and pursue the primary goals of imprisonment and rehabilitation of prisoners. At penitentiaries, assessment reports are not done by the institution itself, but are instead mostly obtained by external, residential psychiatric or psychological experts, who are in most cases not affiliated to a specific academic or scientific institution. Overall, the institutions represent common general forensic-clinical practice in Germany.

1.2 Quality of Criminal Expert Witness Reports

1.2.1 Quality of Risk Assessment Reports

Criminal risk assessment reports about individuals convicted of sexual and/or violence offenses are an increasing issue of public concern (Kury & Obergfell-Fuchs, 2012). In this context, research literature of the last few decades points to a relatively heterogeneous quality of expert witness reports in terms of e.g. insufficient anamnesis, lacking psychometric testing, or inadequate documentation of results (Nowara, 1995a, 1995b; Suhling, 2003; Dahle et al., 2009, 2012; Haarig et al., 2012; Kunzl & Pfäfflin, 2011), which are therefore all addressed in the published minimum requirements for risk assessment reports (see Chapter 1.2.1.1). Numerous references also identify the lack of criteria orientated criminal risk assessment

procedures in expert reports, calling attention to the remaining prevalence and limited accuracy of unstructured, intuitive clinical judgments based only on subjective experience of the expert, not representing scientifically sound expertise. Assessment reports should though not be conducted intuitively on a basis of subjective evaluation criteria but should always follow strictly scientific, transparent, and evidence-based standards.

1.2.1.1 Publication of Minimum Requirements for Risk Assessment Reports

In 2006, specific methodological minimum requirements for risk assessment reports were published (Boetticher et al., 2006), linking empirical scientific evidence to expert practice. These requirements present a scientifically based method and procedure to a standardised individual risk assessment on a basis of empirical knowledge about risk and protective factors for recidivism. Risk assessment reports therefore must record the specifics of the individual biography, the development of delinquency, as well as issues of mental disorders, and background of former crimes. Finally, these data must be included and summarized in an individual theory of delinquency based on a scientific background (Boetticher et al., 2006; 2019).

Empirical evidence to date of how these minimum requirements are put into practice and whether the publication of such minimum requirements has led to greater compliance with these requirements is not yet available. It also seems still unclear, whether the degree of application of these minimum requirements is related to hit rates of prognostic decisions.

In the relevant legal texts on the judicial system, various and sometimes vague formulations of expert´s assignments regarding criminal risk assessment can be found (Dahle, 2010; Boetticher et al. 2019). However, in order to be able to meet the expectations and needs of the recipients, judicial orders are required to explicitly formulate the expert´s assignment in order to clarify which questions should and should not be addressed within risk assessment reports. Simplified, it can be concluded that "the more detailed the formulated assignment, the better the assessment report" (Böhm, 2018, p. 134). In order to ensure more consistent

answers to fundamental prognostic questions by experts despite these different legal requirements and matters of prognostic questions, further specific methodological recommendations for risk assessment reports were published (Boetticher et al., 2019). Besides updating the methodological minimum requirements (Boetticher et al., 2006), concrete fundamental questions to which "even the judicial order should at least be oriented " (p. 555) were formulated from a legal perspective: What is the probability of re-offenses (1)? What will be the nature, frequency, and severity of these crimes (2)? What possible risk-reducing interventions can be taken (3)? What are possible risk-increasing circumstances (4)? The judicial order should "precisely describe the assessments matter and clarify which questions are actually to be answered by the expert; the mere reproduction of the legal text is regularly not sufficient for this purpose" (Boetticher et al., 2006, p. 539). Also, there is no empirical evidence as to which extent these requirements are implemented by judicial orders, nor if or how expert statements answer these fundamental questions in their criminal risk assessments. Furthermore, there is no evidence of the influence of judicial orders on the quality of risk assessment reports. As the recommendations for risk assessment reports reports are also intended to contribute to a dialogue capability between orderers and experts (Nedopil, 2005; Pfister, 2019), it seems relevant to examine to what extent explicit formulations of judicial orders influence the quality of risk assessment reports practice.

1.2.1.2 Superiority of Standardized Criminal Risk Assessment Approaches

Despite the orientation by the publication of minimum requirements, criminal risk assessment reports remain a professional, methodological, and clinical challenge for experts (Rettenberger & Eher, 2016). They show an increasing complexity of differentiated questions, which require the analysis of different data and diagnostic findings and their methodologically sound integration based on valid assessment models (Dahle & Lehmann, 2018). The scientific literature discusses different approaches for criminal risk assessment, which may be basically divided into subjective clinical (or unstructured, intuitive, unguided, or impressionistic),

actuarial (also including statistical, mechanical, or algorithmic) as well as structured professional or clinical-idiographic predictions (e.g., Grove et al., 2010; Meehl, 1954; Nicholls et al., 2013), each with advantages and limitations (Rettenberger, 2018). Subjective clinical judgments are defined as assessments based solely on clinical experience and judgment of assessors using informal and subjective methods, which are predominantly justified by their training, expertise, and professional designations (Grove et al., 2010; Hanson & Morton-Bourgon, 2009; Skeem & Monahan, 2011). This approach is referred to as "unstructured" because of its lack of explicit rules for assessors, which increases its vulnerability to biases and as a consequence its limited reliability and validity (Bengtson & Långström, 2007; Nicholls et al., 2013). Contrarily, highly structured actuarial risk assessment instruments (ARAIs) contain a predetermined list of empirically derived static and/or dynamic risk factors correlating with recidivism and a statistical or algorithmic model to combine the risk factor scores into a total score (Grove et al., 2010; Nicholls et al., 2013), in order to infer empirically determined probability values from those (Dahle, 2005; Hanson, 2009). Checklists or criteria catalogues that serve to clinically structure risk and protective factors can be summarized under the methodological approach of structured clinical assessment instruments (so-called Structured Professional Judgement [SPJ]; von Franqué, 2013). The approaches serve as professional guidelines integrating also the risk-related information derived from a predetermined list of risk (and sometimes protective) factors. In contrast to actuarial instruments, the item values are not added up to an overall value but are used in the sense of a clinical explanatory model, on which basis the individual risk of recidivism is to be structurally assessed. The application of SPJ guidelines is based on an idiographic interpretation of the relevance of each factor and implies a clinical (rather than an algorithmic or statistical) processing of the risk-related data (Nicholls et al., 2013). In a nutshell, a decision based on ARAIs is usually made mechanically according to a fixed algorithm, while in SPJ instruments the final decision is made by a structured assessment

based on personal expertise and the theoretical and empirical knowledge of a professionally educated assessor (Lodewijks et al., 2008). In terms of generations (Andrews et al., 2006; Taxman, 2017), unstructured clinical judgment [UCJ] is considered as the first generation of risk assessment, followed by the second (consisting of static risk factors) and third generations (containing dynamic risk factors) of ARAIs and the SPJ approaches (fourth generation). The combination of actuarial and SPJ instruments with a clinical individual case-based assessment method is often referred to as a - comparatively complex – clinical-idiographic approach (for a detailed discussion, cf. Rettenberger, 2019). It combines actuarial knowledge and structured professional judgments with theoretically sound explanations of the individual behaviours including clinical (e.g., neuropsychological or psychopathological) aspects of the examinee by strictly following scientific standards (Craig & Rettenberger, 2018; Craig et al., 2020).

While the 2nd and 3rd generation actuarial approaches imply a standardized and more empirically orientated approach based on group statistically derived findings, the degree of individual case-based orientation required by legislation (Boetticher et al., 2019; Kröber et al., 2019) increases with the use of SPJ instruments as well as clinical-idiographic approaches. However, standardized risk assessment instruments always provide empirically based probability statements which are an essential aid to decision-making in the required individual case-based assessment. A thorough and systematic risk assessment by statistical-actuarial approaches, with reference to group-statistical average correlations, forms a sound basis on which individual case-related hypotheses regarding possible risk areas can be derived, taking into account the broadest possible empirical knowledge (Dahle, 2010). The use of appropriate instruments serves to reliably determine baseline risk, identify care and treatment goals, and transparently present the course of treatment. Taken together, the use of scientifically sound risk assessment instruments can therefore be considered an integral part of criminal risk assessments as well as for the planning and implementation of the most effective risk

management measures possible and as state of the art in current risk assessment practice (Dahle, 2010; Rettenberger, 2019). The standardized instruments allow an assignment of relevant subgroups of offenders, on the basis of which valid statements about expected recidivism rates in terms of relative and absolute risk can be derived (Eher et al., 2019). However, it does not seem possible to formulate generally applicable guidelines or recommendations for the use of specific risk assessment instruments, as this must always be oriented to the context and the specific question posed (Kröber et al., 2019; Rettenberger, 2019).

When it comes to research regarding the comparison between these different assessment approaches, several studies have underlined the superiority of structured methods (second, third, and fourth generation) and the limited accuracy of UCJ (first generation) not only for criminal risk assessment settings, but for a diverse range of different aspects of human behaviors (e.g., Ægisdóttir et al., 2006; Bengtson & Långström, 2007; Hanson & Morton-Bourgon, 2009). Those findings have been repeatedly and unambiguously confirmed. A number of previously published studies consistently highlighted that unstructured assessments were significantly more susceptible to biases (e.g., Grove et al., 2000; Johansen, 2006; Turgut et al., 2006) along with statistically based and standardized risk assessment instruments showing a significantly higher predictive validity compared to judgments based on clinical intuition or experience (e.g., Rettenberger, 2018; 2019; Viljoen et al., 2021). Intuitive and experience-based risk assessment reports often do not show a predictive validity above chance, independent of data quality and professional experience of experts (Grove & Meehl, 1996; Grove et al., 2000; Hanson & Morton-Bourgon, 2009, Quinsey et al., 2006), and can thus be regarded as partly responsible for the heterogeneous quality in risk assessment reports (Rettenberger & Eher, 2016; Wertz et al., 2018). Particularly for the prediction of sexual or violent recidivism, the structured method is considered as clearly superior compared to unstructured approaches (e.g., Bonta et al., 1998; Heilbrun et al., 2010; Jackson et al.,

2004). Therefore, in the last decades a multitude of ARAIs and SPJ tools for assessing the recidivism risk of different populations and for different settings and outcomes have been developed (e.g., Guy, 2008; Singh et al., 2014; Viljoen et al, 2021) and examined for their psychometric criteria (Rettenberger, 2019). This has contributed decisively to an increasing improvement in the predictive accuracy of risk assessments (Gretenkord, 2013; Rettenberger & Eher, 2012). However, when direct comparisons between ARAIs and SPJ instruments have been carried out no clear and consistent superiority for either has been identified (e.g., Campbell et al., 2009; Douglas et al., 2005; Hanson et al., 2007).

Given this frequently reported superiority of structured risk assessment approaches compared to unstructured and intuitively made judgments and its consequent widespread use in forensic institutions and in expert witness reports (e.g., Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019; Singh et al., 2014), it seems to be surprising at first glance that various aspects in this field of assessment research and practice are still discussed controversially. In this sense, several studies and reviews pointed out the methodological limitations of the existing research (e.g., Dressel & Farid, 2018; Litwack, 2001; Mossman, 1994) and indicated that the accuracy of unstructured assessments might not be as limited as stated (e.g., de Vogel et al., 2004; Lin et al., 2020; Melton et al., 2018). For example, a recent umbrella review of systematic reviews comparing structured and unstructured risk assessment methods reported that “although research is generally consistent in reporting that risk assessment tools are superior to UCJ, studies used to support this statement showed serious problems in terms of risk for bias and lack of direct comparison” (Viljoen et al., 2021, p. 92). The authors showed, for example, that almost the entire state of research was conducted decades ago and nearly two thirds of the primary studies included in most reviews were published in the 1980s or earlier. Furthermore, only a few studies compared directly structured and unstructured assessment approaches and did not examine whether the predictive validity differed significantly but based their conclusions on a visual inspection of the results. Several studies did not focus particularly on

risk assessment instruments and did not provide direct head-to-head comparisons between ARAIs or SPJ tools and UCJs. Given these research desiderata, further empirical studies which directly compare unstructured with structured risk assessment methods are highly recommended (Viljoen et al., 2021).

1.2.1.3 Degree of Standardization in Criminal Risk Assessments in Professional Practice

On the basis of the ongoing debate on limited accuracy of unstructured clinical judgments and in spite of a variety of scientifically based methods of criminal risk assessment (Rettenberger & von Franque, 2013), the intuitive and experience-based approach seems to remain prevalent (Haubner-Mclean et al., 2014). It therefore remains unclear, to what extent of standardization criminal risk assessments are actually conducted in present professional practice. Despite numerous research papers on the reliability and (predictive) validity of standardized criminal risk assessments, comparatively few studies can be found that dealt with the actual use of actuarial and clinical-structured risk assessment instruments or psychometric tests in real practice. There is also relatively little evidence yet about how standardised criminal risk is finally communicated in risk assessment reports to decision makers (de Vogel et al., 2020).

While at least some studies are available for the international, primarily Anglo-American area (e.g., Archer et al., 2006; Singh et al., 2014; Viljoen et al., 2010), only a few empirical studies have so far been conducted for the German-speaking area on the form in which criminal risk assessments are made in professional practice. There is particularly little evidence about the risk assessment practice, especially of external experts. A survey study on criminal risk assessment practice in Germany (Rettenberger at al., 2017) showed that, according to their own statements, psychological and psychiatric experts use standardized risk assessment instruments in the majority of cases. Standardized risk assessment instruments were used in over 50% of all assessments, and in the period of the twelve months prior to the survey, practitioners reported using a risk assessment instrument in 65% of cases. The group

of psychological experts used risk assessment instruments significantly more often than their psychiatric colleagues. The most frequently cited instruments were the Psychopathy Checklist-Revised (PCL-R; Hare, 2003), the Historical Clinical Risk Management-20 (HCR-20; Müller-Isberner, et al., 1998), the Forensic Operationalized Therapy Risk Evaluation System (FOTRES; Rossegger et al., 2011), and the Violence Risk Appraisal Guide (VRAG; Quinsey et al., 2006). According to the survey, the use of actuarial and clinical-structured instruments in this context not only served to assess future anticipated offending but was also used for treatment planning and to review the course of treatment. A survey of department heads and specialized services on intramural psychodiagnostic practice in social therapy facilities of penitentiaries (Etzler & Rettenberger, 2019) revealed that initial diagnostics were more standardized than ongoing and final diagnostic examinations. A total of 23 different risk assessment instruments were mentioned, in descending order of frequency of use: the HCR-20, the PCL-R, the Level of Service Inventory-Revised (LSI-R; Dahle et al., 2012), the Sexual Violence Risk-20 (SVR-20; Müller-Ibserner et al., 2000), the VRAG, the Static-99 (Harris et al., 2003), the Sex Offender Risk Appraisal Guide (SORAG; Quinsey et al., 2006), the so-called Dittmann List (Dittmann, 2000), and the Stable-2007 (Matthes & Rettenberger, 2008). These survey results also suggested that criminal risk assessment practice in social therapy settings is relatively heterogeneous. In another survey study on criminal risk assessment practice in outpatient aftercare for individuals convicted or charged of sexual offenses in Germany (Gregório Hertz et al., 2019), more than three-quarters of participating institutions reported using criminal risk assessment instruments. The most frequently cited instruments were the Static-99, the Stable-2007, the PCL-R, the LSI-R, the HCR-20, the VRAG, the Dittmann List, and the Structured Assessment of Protective Factors for Violence Risk (SAPROF; de Vogel et al., 2011; Yoon et al., 2013), again showing the heterogeneity of the use of standardized risk assessment instruments.

Furthermore, the standardised presentation and communication of risk is essential for effective risk management and should ideally be transparent, comprehensible and unambiguous (Boetticher et al., 2019). Basically, a distinction can be made between a nominal and quantitative risk communication. Nominal risk communication involves a dichotomous or categorical weighting of risk, for example in the form of "low," "moderate," or "high" risk, which, however, may not be subject to any clear assignment to recidivism probabilities and thus to a significant extent to intuitive subjective evaluation criteria. In numerical communication, probability or frequency information is expressed in the form of category-specific absolute recidivism rates for defined periods of time, usually requiring a reference value (e.g., in the form of recidivism base rates) for a content classification of the quantitative information (de Vogel et al., 2020; Eher et al., 2019; Gretenkord, 2013; Nedopil, 2005). The "5-Category Model for Sexual Offenders" provides a merging of relative and absolute risk based on transparent, comprehensible, combined nominal and quantitative risk communication orientated to the recidivism baseline rate (Eher et al., 2019), which was initially developed for the Static-99, but in the future will pe provided for other instruments. Other authors/researchers also highlight the benefits of a combined, integrative risk communication consisting of categories and probability scores (e.g., de Vogel et al., 2020). According to the current recommendations for risk assessment reports (Boetticher et al., 2019), the risk of recidivism posed by an assessed person must also be sufficiently specified: "(...) in particular, the probability of recidivism and nature of offense must be substantiated" (p. 559). Different international studies (Heilbrun et al., 2016; Singh et al., 2014) suggested a clear preference of forensic experts for a categorical risk communication over a dichotomous or probabilistic form of communication. A survey study of risk assessment practice in Germany (Rettenberger at al., 2017) also showed that the preferred form of risk communication referred to a categorical assessment (in terms of the SPJ approach of "low," "moderate," or "high" risk).

Because a structured clinical approach is also indispensable in forensic assessments, clinical conclusions about relevant aspects of personality, cognitive functions, intelligence, or mental disorders of examinees in assessment reports should also follow strict scientific standards, also including the use of standardized psychological measures based on scientific theory and technique. Due to psychometric characteristics of psychological tests, relevant aspects can be measured more objectively, reliably, and validly by implementing a distinct degree of standardisation and formalisation, especially in terms of standard values and norm data. Besides clinical interviews, risk assessment instruments and file information, experts therefore consequently base their assessments on empirically supported psychological tests to gain psychometrically sound foundation of reports (American Psychological Association, 2013; McLaughlin & Kan, 2014; Miller, 2013). Hence, psychological assessment is a defining area of practice, training, and research for professional psychologists, as a large majority believe psychological assessment is a valuable aid in assisting diagnostic and treatment decisions or in screening for cognitive or neuropsychological deficits (Wright et al., 2017).

Therefore, “Specialty Guidelines for Forensic Psychology” (American Psychological Association, 2013), a “Practice Guideline for the Forensic Assessment” by the American Academy of Psychiatry and the Law (Glancy et al., 2015), or “Best Practices in Forensic Mental Health Assessment” (Heilbrun et al., 2009) were developed and published internationally. A German assessment system of psychological tests also provides additional evaluation criteria for usage of psychological testing, also including description of type, content and diagnostical purpose as well as norm sample, reliability, and validity of tests (Diagnostik- und Testkuratorium, 2018). Furthermore, also the methodological minimum requirements for criminal responsibility (Boetticher et al., 2007) and risk assessment reports (Boetticher et al., 2006; 2019) included recommendations for usage of psychological testing. According to these requirements, practical use and selection of psychological tests are based

on psychometric quality criteria validated by empirical studies. Tests also have to meet requirements according to adequate norm samples of the German-speaking area.

Psychological test expertise is either part of a forensic expert report or forensic psychologists are asked to conduct an additional specific psychological test report for other mental health professionals like, for example, forensic psychiatrists (Rieger & Stadtland, 2013; Schneider et al., 2015; Wertz & Tippelt, 2019). Generally, the frequency of the usage of psychological testing within court statements is increasing (Habermeyer, 2008). Although psychological testing seems to have a general acceptance within the forensic field in Germany (Habermeyer, 2008; Schneider et al., 2015), empirical findings about the usage and application of tests and especially about the consideration of forensic psychological testing in the final judgments of risk or criminal responsibility of individuals charged or convicted of offenses is still sparse. A German study found an increased usage of test instruments of 55% in 227 forensic evaluations regarding the preventive detention in comparison to reported 20% in a period of ten years (Habermeyer, 2008; Kinzig, 1997a, 1997b). A survey of the diagnostic practice in social therapy units in the German prison system (Etzler & Rettenberger, 2019) showed that psychological testing plays a key role in planning and monitoring treatment but at the same time indicates a relatively heterogenous diagnostic practice in these institutions. Another survey described the outpatient aftercare of individuals convicted of sexual offenses after release from prison in Germany, indicating a regular use of standardized diagnostic tests, mostly multidimensional personality instruments (Gregório Hertz et al., 2019).

A range of studies has reported that psychological testing is applied in a variety of different types of forensic evaluations, including risk and criminal responsibility assessments (Archer et al., 2006; Borum & Grisso, 1995; Lally, 2003; McLaughlin & Kan, 2014; Neal et al., 2019; Varela & Conroy, 2012; Viljoen et al., 2010; Watkins et al., 1995; Wright et al., 2017; Zapf et al., 2004). Different studies indicated that within forensic psychology and psychiatry, psychological testing seems to be a commonly used source of information in

expert reports (e.g., Cutler & Kovera, 2011; McLaughlin & Kan, 2014; Neal & Grisso, 2014; Serafim et al., 2015; Wright et al., 2017). Previously published surveys reported that among forensic psychologists, approximately 30% of the working time in forensic-clinical practice is spent in psychological testing (Archer et al., 2006). A more recent survey reported an average of 24% of direct practice time conducting any psychological assessment, while participants working in a forensic setting spent a significantly greater percentage of their time with psychological testing than those in non-forensic institutions (Wright et al., 2017). Although surveys consistently identified psychological testing as an important component of forensic evaluations, frequency, nature, and concrete application practice of tests and therefore degree of formalisation vary significantly by setting, legal question, and evaluation focus (Archer et al., 2006; Fuger et al., 2014; Lally, 2003; McLaughlin & Kan, 2014; Neal, 2018).

Due to survey data studying the frequency and acceptability of psychological tests in different areas of forensic psychology (Archer et al., 2006; Boccaccini & Brodsky, 1999; De Clerq & Vander Laenen, 2019; Lally, 2003; McLaughlin & Kan, 2014; Neal, 2018; Neal et al., 2019; Neal & Grisso, 2014; Viljoen et al., 2010; Wright et al., 2017), multiscale personality tests (like the Minnesota Multiphase Personality Inventory-2 [MMPI-2; Engel et al., 2000], the Personality Assessment Inventory [PAI; Morey, 1991] or the Revised NEO Personality Inventory [NEO-PI-R; Costa & McCrae, 2008]) were generally mentioned most frequently. Consistent with the results of previous surveys, findings underline the continuing popularity of traditional clinical assessment instruments in forensic psychology, such as the MMPI-2 (Archer et al., 2006; Lally, 2003; Mclaughlin & Kan, 2014; Neal et al., 2019; Wright et al., 2017). Rated as suitable to examine the mental state at the offence in terms of cognitive and intellectual functions, various forms of the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997) were recommended consistently by survey respondents, followed by other performance tests (Archer et al., 2006; Boccaccini & Brodsky, 1999; Lally, 2003; Neal & Grisso, 2014; Viljoen et al., 2010; Wright et al., 2017). Another one of the most commonly

used group of assessment instruments were the symptom specific measures or inventories like the Beck Depression Inventory (BDI; Beck et al., 1979), as reported by a recent survey (Wright et al., 2017).

Contrary to the findings of Lally (2003) and McLaughlin and Kan (2014) - indicating that performance-based personality tests, typically classified as projective measurements (Meyer & Kurtz, 2006; Viglione & Rivera, 2003), were neither accepted nor were the least often used in practice—and despite science-based criteria for the admissibility, the predominant number of surveys and examinations of court-ordered expert reports suggest that multiple performance-based personality tests (like the Rorschach test [Rorschach, 1927] or the Thematic Apperception Test [Morgan & Murray, 1935]) continue to be used by a substantial number of psychologists (Archer et al., 2006; De Clerq & Vander Laenen, 2019; Neal et al., 2019; Viljoen et al., 2010). Taken together, international surveys reported that most forensic evaluations are using multiple psychological tests, noting the extreme variety regarding the number of tools used (De Clerq & Vander Laenen, 2019; Fuger et al., 2014; Lees-Haley et al., 1996; McLaughlin & Kan, 2014; Neal & Grisso, 2014; Warren et al., 2006; Wright et al., 2017).

However, despite these studies, other authors have criticized the fact that the knowledge of forensic psychological testing in practice is still lacking compared with general clinical practice (Lally, 2003; Neal et al., 2019). This means that, on the one hand, there are some comparatively frequently used and relatively popular measures like the MMPI-2 or the Wechsler Intelligence scales but, on the other hand, there is no consensus about psychological test usage and the degree of formalisation in forensic evaluations (Archer et al., 2006; Golden & Lashley, 2014; Gowensmith & McCallum, 2019; Lally, 2003; Richards et al., 2015). Even though studies exist that survey forensic psychologists about the tests they are using regularly, the knowledge about forensic psychological testing in forensic-clinical practice is still sparse compared to other (i.e., non-forensic) clinical settings (Archer et al., 2006; Lally, 2003). For

example, there is still little empirical data about reporting and consideration of psychological test results in forensic court reports.

1.2.1.4 Excursus: Risk Assessment in Criminal Responsibility Assessment Reports

In cases where ordering of mandatory treatment is being considered, experts are also regularly asked for a criminal risk assessment in addition to the assessment of the prerequisites of the suspended or diminished criminal responsibility (Müller & Nedopil, 2017). A placement in a forensic psychiatric hospital may only be ordered if it is related to the significance of the offenses committed and those that can be further expected as well as the degree of dangerousness posed by the examinee (Kammeier, 2018). Criminal risk assessments are not only linked to the existence of a crime-relevant mental disorder (article 63 of the German penal code) or a crime-relevant tendency towards intoxication (article 64 of the German penal code) and significant offenses (article 66 of the German penal code). Experts should also provide explanations on the probability, nature, and severity of future offenses as well as risk-altering measures and circumstances (Boetticher et al., 2019). In this respect, judicial questions about the prerequisites of a mandatory treatment require the expert to deal with the risk of recidivism in addition to criminal responsibility. So far, it hardly seems to be empirically recorded how the risk assessment – i.e., the judicial question regarding the risk of significant offenses expected in the future as a prerequisite for the ordering of a mandatory treatment – is implemented in criminal responsibility reports.

As shown in current studies, the overturning of sentences due to defective criminal responsibility assessment reports are oftentimes based on risk assessments in the context of the articles 63 and 64 of the German penal code. Since, according to case law, the decisive criterion is a high degree of probability for the commission of significantly unlawful acts, it is not sufficient from an expert point of view to assess whether such acts are "to be expected," "possible," "not improbable," or even "not excludable" (Mosbacher, 2020, p. 446ff). The prerequisites of mandatory treatment primarily deal with one’s individual dangerousness due

to mental disorders, which can only be assumed if there is a justified expectation of significantly unlawful acts.

The increasing use of standardized risk assessment instruments has been noted in the expert context (Dahle & Lehmann, 2018; Leygraf, 2015; Müller & Nedopil, 2017; Rettenberger, 2018; Rettenberger & von Franqué, 2013), in the context of social therapeutic institutions (Etzler & Rettenberger, 2019), in the practice of outpatient aftercare for individuals convicted of sexual offenses (Gregório Hertz et al., 2019), as well as in the practices of granting privileges in forensic psychiatric hospitals (Sklenarova et al., 2020). The added value of standardized risk assessment instruments is also emphasized in the methodological minimum requirements or recommendations for risk assessment reports (Boetticher et al., 2019; Kroeber et al., 2019). However, yet no empirical data are available on the extent to which this development can also be found in criminal responsibility assessment reports.

1.2.2 Quality of Criminal Responsibility Assessment Reports

The criminal responsibility assessment is of particular importance due to the considerable consequences for the individuals assessed in the criminal proceedings. It has a decisive influence on sentencing decisions regarding placement in a forensic hospital and the duration of the deprivation of liberty (Müller & Nedopil, 2017; Prüter-Schwarte et al. 2019; Verrel, 1995). Criminal responsibility assessment reports frequently attract public interest and are regularly the subject of media discussions (Dahle et al., 2012; Kury & Obergfell-Fuchs 2012; Müller & Nedopil2017; Verrel, 2015). The scientific literature points to a heterogeneous quality of expert witness reports, which, in addition to formal deficiencies, also refers to aspects of content (Dahle et al. 2012; Fegert et al., 2003; Kunzl et al., 2009; Kunzl & Pfäfflin 2011; Schläfke et al., 2006; Schnoor, 2009). Regarding the legal side, it is also not uncommon for overturned sentences to be justified by errors in the criminal responsibility assessment or the assessment of the prerequisites of mandatory treatment in a forensic mental

hospital (Mosbacher, 2020, p. 446). As a result of the ongoing discussion about the quality of expert witness reports and quality assurance, methodological minimum requirements for criminal responsibility assessments were published in 2007 (Boetticher et al., 2007). Empirical evidence on the extent to which these methodological minimum requirements are implemented in practice is still scarce. An empirical study conducted in Germany after the publication of the methodological minimum requirements still found urgent potential for improvement in almost 200 analysed criminal responsibility assessment reports (e.g., Prüter-Schwarte et al., 2019). While the formal methodological minimum requirements were largely implemented, "considerable weaknesses [...] were found in areas essential for the assessment of criminal responsibility in the assessment reports examined, which question the results of the assessments" (Prüter-Schwarte et al., 2019, p. 207). Another study of 50 German criminal responsibility reports conducted in 2017 still showed formal inadequacies and deficiencies in content (Stübner et al., 2018). The extent to which the publication of the minimum requirements led to an improvement in the quality of criminal responsibility assessments over time has not yet been examined. Although the methodological minimum requirements for criminal responsibility assessment reports were primarily intended for forensic experts, they should also provide support for judges, public prosecutors and defense attorneys. They should help assess the validity of the reports and uncover dubious expertise and contradictions between experts (Boetticher et al., 2007). The minimum requirements are intended, among other things, to contribute to a dialogue capability between orderers and experts (Pfister, 2019). Therefore, as a first step, it seems relevant in terms of quality assurance to consider how transparently the prerequisites of the suspended or diminished criminal responsibility are discussed in assessment reports and, second, to examine the extent to which expert recommendations are adopted by the courts and reported in the verdict. Previous studies (Schläfke et al., 2006; Schnoor, 2009) suggest that the courts largely adopt the results of the expert reports on criminal responsibility, and usually all recommendations are incorporated

verbatim into the verdict without further discussion. A critical examination of the expert recommendations or a discussion of these was only rarely present (Schläfke et al., 2006; Schnoor, 2009; Verrel, 2015). The review of judicial adoption behaviour after the publication of the methodological minimum requirements showed a relation to the proportion of fully met minimum criteria at a similarly high adoption rate. At the same time, however, there is still a predominantly formulaic to absent judicial discussion of the expert reports (Schöttle et al., 2013).

1.3 Aim and Subject of the Present Thesis

Given that criminal risk assessment decisions entail far-reaching freedom-related consequences for individuals (Rettenberger, 2018; 2019) regarding (the duration of) the deprivation of liberty (Müller & Nedopil, 2017; Prüter-Schwarte et al., 2019; Verrel 1995; 2015), research literature pointing to a heterogeneous quality of expert witness reports provided the starting point for the research presented in this thesis. As most scientific references are based on subjective impressions of experienced expert witnesses or presentations of individual cases without profound empirically sound foundation, the aim of this thesis was to systematically examine the quality of criminal risk assessment reports in German practice retrospectively. Therefore, the thesis is based on an iterative research project of six separate empirical studies, retrospectively analysing more than 1.000 risk and criminal responsibility assessment reports from different institutions representing common general practice in Germany.

The first study (Chapter 2; Wertz, Kury & Rettenberger, 2018) was conducted in order to systematically examine the quality of criminal risk assessment reports with special regard to the published methodological minimum requirements (Boetticher et al., 2006). Since the publication of methodological minimum requirements for risk assessment reports of an interdisciplinary working group in 2006 in Germany, there has been no empirical investigation into whether these minimum requirements are put into clinical practice and

whether the publication of such minimum requirements has led to an improvement in the quality of risk assessment reports. It also seems still unclear, whether the degree of application of these minimum requirements is related to hit rates of prognostic decisions. The first study remedied this research desideratum by providing an empirical analysis of the application of these methodological minimum requirements for risk assessment reports of individuals convicted of sexual and/or violent offenses from the penitentiary in Freiburg and the Department of Forensic Psychiatry of the University Hospital Munich before (from 1999 to 2002) and after (from 2008 to 2011) the publication in 2006. Additionally, actual re-offenses according to the Federal Central Register (June 2016) were used to take a closer look at the relation of the degree of application of the above-mentioned minimum requirements and hit rates of prognostic judgements.

According to the current recommendations regarding criminal risk assessments, judicial orders are required to explicitly formulate the expert's assignment, to describe the assessment's matter, and to clarify the exact questions the expert is supposed to answer. The assessment should therefore be oriented along four fundamental prognostic questions. Accordingly, the second question addressed in this thesis was to which extent these requirements are implemented by judicial orders, nor if or how expert statements answer these fundamental questions in their criminal risk assessments. To empirically examine the research question, judicial orders and the answers to the prognostic questions in the risk assessment reports were systematically gathered within the second study (Chapter 3, Wertz et al., 2020). It was examined whether a stronger orientation to the methodological minimum requirements can be found over time and to what extent there is a relation between judicial orders and their answers in risk assessment reports.

As the research literature not just only points to a heterogeneous quality of criminal risk but also of criminal responsibility assessments in practice, minimum requirements were also implemented for criminal responsibility assessment reports. Empirical evidence on the

extent to which these methodological minimum requirements are implemented in practice is still scarce. In cases where ordering of mandatory treatment is being considered, experts are regularly asked for a criminal risk assessment in addition to the assessment of the prerequisites of the suspended or diminished criminal responsibility (Müller & Nedopil, 2017). So far, it hardly seems to be empirically recorded how the risk assessment is implemented in criminal responsibility reports and if the development of an increasing use of standardised risk assessment instruments can also be found in criminal responsibility compared to risk assessment reports. As an excursus, the aim of the third study (Chapter 4, Wertz et al., 2021) was therefore to examine whether the publication of methodological minimum requirements (Boetticher et al., 2007) has led to an increased implementation of these in criminal responsibility assessment reports in clinical practice, how the criminal risk assessment required by the text of the law as a necessary prerequisite for mandatory treatment according to articles 63, 64, and 66 of the German penal code is implemented in criminal responsibility assessment reports, whether published recommendations for criminal risk assessment reports have led to a stronger orientation towards these in the risk assessment in criminal responsibility assessment reports, and the form in which the expert findings are considered in in judicial verdicts.

Within the fourth (Chapter 5, Wertz et al., 2022) and fifth study (Chapter 6, Wertz & Rettenberger, 2021), the standardization of criminal risk and responsibility assessment reports was addressed in terms of the use of psychometric tests (Chapter 5) and structured criminal risk assessment instruments (Chapter 6) in professional practice. Despite numerous research papers on the reliability and (predictive) validity of standardized psychometric tests and criminal risk assessments, comparatively few studies can be found that dealt with the actual use of those instruments or psychometric tests in real practice. While in previous studies the survey data were usually based on the self-reports of clinicians (e.g., Archer et al., 2006; Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019; Lally, 2003; Rettenberger at al.,

2017; Wright et al., 2017), in the present investigation a comprehensive amount of court reports about risk assessment and criminal responsibility cases were empirically analysed. Therefore, study objectives were to examine the number, frequency, content, and type of tests and risk assessment instruments used in assessment reports, and the reporting practice and influence of test results on assessment reports (especially with regard to the final judgments). Whether there are differences in the use of structured risk assessment instruments depending on report- and examinee-related characteristics as well as how risk communication is carried out in such expert reports in which standardized risk assessments instruments are used, also remain open questions so far and were therefore addressed.

On the basis of the ongoing debate on limited accuracy of unstructured clinical judgments compared to structured approaches, the sixth study (Chapter 7, Wertz et al., in press) finally directly compared the predictive accuracy of unstructured as well as different structured risk assessment approaches derived from a comprehensive sample of German risk assessment reports about individuals convicted of violent and/or sexual offenses. More precisely, the first study objectives were to identify and classify the applied assessment approaches in the risk assessment reports due to their degree of structuring the assessment process. Second, we aimed to compare the predictive accuracy of unstructured and structured risk assessment approaches (i.e., unstructured clinical judgment [UCJ], ARAIs [in terms of the revised version of the Violence Risk Appraisal Guide, VRAG-R; Rice et al., 2013], SPJ instruments, and combinations of ARAIs-/SPJ-based risk assessments). In order to examine the stability and generalizability of the results, different follow-up periods (5-year and in total) and offence types (general, nonviolent, violent, general sexual, and sexual contact recidivism) have been used. Based on the existing state of research we hypothesized that structured risk assessment methods would outperform unstructured risk assessment procedures in predicting general, sexual (contact) and violent recidivism. However, between the different types of structured risk assessment methods we did not expect significant

differences with the exception that we would expect the highest predictive accuracy for the combination of different structured assessment approaches (i.e., the combined use of SPJ and ARAIs).

2. The Application of Methodological Minimum Requirements for Risk Assessment Reports in Clinical Practice - An Empirical Validation Using Officially Registered Reoffenses[2]

Abstract

Risk assessment reports regarding individuals convicted of sexual and/or violent offenses are an issue of increasing public and socio-political concern, and their methodological quality is regularly and intensively discussed in research and practice. The review of research literature distinctly shows several indicators for a heterogeneous quality of these expert witness reports in practice. Since the publication of methodological minimum requirements for risk assessment reports of an interdisciplinary working group in 2006 in Germany, there has been no empirical investigation into whether these requirements are put into clinical practice and whether the quality of risk assessment reports has increased since then. The application of these methodological minimum requirements for risk assessment reports of individuals convicted of sexual and violent offenses from the penitentiary in Freiburg and the Department of Forensic Psychiatry of the University Hospital Munich (N = 502) from 1999 to 2002 and from 2008 to 2011 was analysed on the basis of a questionnaire gathering the professional background and institutional affiliation of expert witnesses, the final prognostic judgement, the central prognostic issue, the index offence, and whether the report was written before or after the publication of the methodological minimum requirements for risk assessment reports. Based on this, the positively directed risk assessments were validated by actual re-offenses according to the Federal Central Register (June 2016) and were related to the degree of application of the above-mentioned minimum

[2] Paper published as Wertz, M., Kury, H. & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [The application of methodological minimum requirements for risk assessment reports in clinical practice - An empirical validation using officially registered reoffenses]. Forensische Psychiatrie, Psychologie, Kriminologie, 12(1), 51 – 60.

requirements. The results showed that there is a slight increase in the quality of the analysed risk assessment reports which can be related to the publication of the minimum requirements. Furthermore, the consideration of the minimum requirements in assessment reports was positively correlated with the hit rates of corresponding prognostic judgements according to officially registered reoffenses. Taken together, the results indicate the need of further discussions about quality management in the field of forensic psychiatry and psychology.

Keywords: Quality (assurance), risk assessment reports, methodological minimum requirements, violent offenders, sexual offenders, recidivism rates

2.1 Quality of Criminal Risk Assessment Practice

Criminal risk assessment reports about individuals of sexual and/or violent offenses are an increasing issue of public concern (Kury & Obergfell-Fuchs, 2012). In this context, research literature of the last few decades points to a relatively heterogeneous quality of expert witness reports (Nowara, 1995a, 1995b; Suhling, 2003; Dahle et al., 2009, 2012; Haarig et al., 2012; Kunzl & Pfäfflin, 2011). In 2006, specific methodological minimum requirements for risk assessment reports were published (Boetticher et al., 2006), linking empirical scientific evidence to expert practice. Empirical evidence of how these minimum requirements are put into practice and whether the publication of such minimum requirements has led to an improvement in the quality of risk assessment reports is not yet available. It also seems still unclear, whether the degree of application of these minimum requirements is related to hit rates of prognostic decisions. In an iterative research project of three independent empirical studies, conducted risk assessment reports were retrospectively analysed regarding quality in terms of application of the minimum requirements over time (Kury et al., 2009; Kury & Adams, 2010; Wertz, 2016). Initial empirical results were provided by Riegl (2007) and Adams (2009; Kury Adams, 2010). They empirically examined the quality of 133 risk assessment reports from external experts of the penitentiary in Freiburg

(Riegl, 2007) and 339 risk assessment reports of penitentiaries in Bruchsal and Freiburg (Kury & Adams, 2010) using a developed questionnaire. Results showed - in addition to a quite heterogenous quality - a significant improvement in the quality of risk assessment reports since the amendment of the “Law to Combat Sexual Offenses and Other Dangerous Crimes" according to article 454 in conjunction with article 66 of the German Panel Code.

2.2 Method

To systematically examine the quality of criminal risk assessment reports with special regard to the methodological minimum requirements (Boetticher et al., 2006), expert reports about individuals convicted of sexual and violent offenses were analysed. Therefore, a questionnaire developed in accordance with previous studies (Kury & Adams, 2010; Riegl, 2007) was used. Results were also embedded in the context of the previous results of the research project. Accordingly, all risk assessment reports of individuals convicted of sexual and violent offenses from the penitentiary in Freiburg, which have not been considered so far, and the archived expert reports of the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig-Maximilians University of Munich, were examined within this study. Thus, a holistic overview of the cohorts from 1999 to 2002 and from 2008 to 2011 within the respective institutions could be gathered (Wertz & Kury, 2017a; 2017b). In addition to the descriptive statistics of experts’ characteristics, examinees, and report-related variables, application of the methodological minimum requirements was assessed. The examination furthermore focused on the professional background and institutional affiliation of expert witnesses, on the final prediction judgement, the central prognostic issue, the index offence, and the question whether the report was written before or after the publication of the methodological minimum requirements. Based on this information, the positively directed risk assessments leading to release or granting privileges were validated by actual re-offenses according to the Federal Central Register (as of June 2016) and were related to the degree of application of the above-mentioned minimum

requirements. Therefore, the Federal Central Register excerpts of all examinees (N = 381) were requested from the Federal Office of Justice in Bonn, whereby n = 19 reports had to be excluded from the evaluation due to deaths or incorrect data of examinees.

2.2.1 Sample

The final sample consisted of n = 294 external risk assessment reports on n = 199 inmates from the penitentiary in Freiburg and n = 208 reports (a total of n = 182 different examinees) from the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig-Maximilian University in Munich regarding the cohorts of 1999 to 2002 and of 2008 to 2011 (N = 502). The assessment reports were ordered by various judicial parties to the proceedings, like local or district courts, courts for the execution of prison, higher regional courts, public prosecutors, district hospitals, the Ministry of Justice of Baden-Württemberg, the Bavarian Court of Justice, penitentiaries, and private law firms. In the university department, a total of n = 17 different department-affiliated experts were involved in the conduction of assessment reports, whereas in the penitentiary in Freiburg a total of n = 71 external experts from independent expert practices, forensic psychiatric hospitals, psychological institutes, or psychotherapeutic and psychiatric hospitals or care centres were consulted. On average, the 17 experts of the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig-Maximilian University of Munich conducted M = 12.2 assessment reports (ranging from n = 1 to n = 98), and the 71 experts of the penitentiary provided an average number of M = 4.1 reports (with a range from 1 to 34 reports). At the Department of Forensic Psychiatry, reports were conducted almost exclusively by psychiatric experts (n = 16, 94.2%). In the penitentiary in Freiburg external psychologists (n = 14, 19.7%), psychiatrists (n = 49, 69.0%), and experts of both professions (n = 8, 11.3%) were asked to assess examinees. Thus, 376 of original 500 assessment reports were provided by medical experts, 100 by psychological experts, and 26 by experts of both professions. A total of 85.0% of the experts were male; in the penitentiary

in Freiburg even more than 90.0%, and in the university department in Munich more than one-third of reports were provided by female experts. With regard to the underlying index offenses, it was found that violent offenses accounted for just under half (*n* = 229, 45.6%) and sexual offenses for just under one-third of all offenses (*n* = 140, 27.9%); homicide and manslaughter offenses (n = 91, 18.1%) as well as a combined occurrence of violent and sexual offenses (n = 42, 8.4%) were also mentioned several times (Kury &Wertz, 2017). Four of the 381 examinees were female; the mean age was *M* = 43.8 years (*SD* = 12.5), and 74.5% had a basic or intermediate school leaving certificate. 11.6% of the examinees had no educational qualification.

In the present study, negatively directed assessments required a higher average of examination time (in weeks) according to the order and date of conduction stated in the expert report (*M* = 16.3, *SD* = 9.8) than positively directed assessments (*M* = 12.7, *SD* = 7.1). In total, the number of total examination days could be extracted from *n* = 471 expert reports (93.8%); a more precise documentation was found in only *n* = 133 expert reports (26.5%), stating hours or minutes. The average regarding the total time of examination was *M* = 2.5 days (*SD* = 1.1) and *M* = 401 minutes (*SD* = 144.7). The maximum duration regarding the total time of examination was 7 days and 900 minutes, respectively. Time of examination in minutes was significantly higher after the publication of minimum requirements (*M* = 421.5, *SD* = 151.9) and in the Department of Forensic Psychiatry of the University Hospital of Munich (*M* = 475.6, *SD* = 180.6) than in the period before 2006 (*M* = 361.0, *SD* = 121.3; *t*(131) = -2.32, *p* = .022) and in the penitentiary in Freiburg (*M* = 375.4, *SD* = 120.9, *t*(43.584) = -3.01, *p* = .004), (Kury & Wertz, 2017).

The number of additional examiners from other professions called in by the principal assessor during the review varied. In *n* = 46 cases, i.e., in 15.6% of all assessments of the penitentiary in Freiburg, an additional examiner was consulted. In 17 cases, the assessment involved an additional psychological testing, in eleven cases an additional exploration, and in

18 cases the additional examiner was assigned to both tasks. Altogether, eleven psychiatrists and 35 psychologists were consulted. In the 208 assessment reports of the university department in Munich, an additional examiner was involved in $n = 61$ cases and two additional examiners in $n = 107$ cases (in total in 80.8% of the cases). In 40 assessments, the task of the additional examiners was psychological testing and was carried out by psychologists, in 24 cases additional exploration by a sociologist, and in 104 assessments both an additional exploration by a sociologist and a psychological test examination by a psychologist was conducted.

Regarding the use of further psychological test diagnostics and risk assessment instruments, 140 reports of the penitentiary in Freiburg lacked information regarding the use of psychological tests or risk instruments, with incomplete documentation in 106 cases and comprehensive documentation in 48 reports. In 59.8% of the cases, test description regarding the type of test as well as information on the psychometric quality criteria was completely missing. Unclear or insufficient documentation was found in 25.2% of the cases, and methodologically sufficient documentation was found in only 15.0%. Referring to the Department of Forensic Psychiatry at the University Hospital of Munich, no information could be extracted in 41 reports, unclear or insufficient documentation was present in 103 cases, and methodologically sufficient documentation was found in 64 reports. The test description was missing in 35.1% of the expert reports, incomplete documentation was found in 58.2% and comprehensive documentation in 6.7% of the cases. However, it should be mentioned, that in many cases an additional report on psychological testing existed but was not accessible. A total of 81 psychological tests and risk assessment instruments was applicated 1634 times, whereby 1087 (66.5%) applications were provided by psychologists and 557 by psychiatrists (33.5%). Intelligence tests, personality tests, and prognostic instruments were used most frequently.

2.2.2 Survey instrument

The survey instrument used (Kury et al., 2009) was based on the minimum requirements for risk assessment reports (Boetticher et al., 2006) and the relevant research literature (Dahle, 2005a, 2005b; Nedopil, 2005; Kröber, 2006). It comprised a total of 312 items, of which 219 were rated on a three-level scale with the characteristics "1 = no information/absent," "2 = incomplete/unclear," and "3 = comprehensively documented." Ninety-three additional items recorded qualitative information regarding characteristics of the experts, the examiners, and report-related variables. If an item could not be evaluated due to comprehensible explanations in the reports and was documented as such an item, it was rated by the highest score. In addition to the overall score for the quality of the reports, further sub scores were determined for differentiated analysis regarding content and form, the procurement and handling of examinee-related information, the methodology, and the final conclusion of the reports (cf. Table 1), whereby the examinee-related information and the methodology were weighted more severely (Kury et al., 2009). Overall, the total score ranged from 140.0 to 436.5 points, where the lowest score meant no consideration of even a single minimum requirement and the maximum value corresponded to an expert report that fully implemented any minimum requirements. The used questionnaire by Riegl (2007) is a non-validated instrument, but the reliability criterion was met by calculating the interrater reliability. Thus, a correlation measure evaluating the strength of the interrater reliability was used to guarantee comparability with the previous studies using the same survey instrument. For this purpose, a random selection of 10.0% of the total number of assessments (Leonhart, 2004) was rated independently by an in-depth and intensively trained co-rater of psychological profession. The ratings of the double-rated expert reports were included in the final statistical analysis (Wertz & Kury, 2017a, 2017b). In accordance to the results of the two previous studies, $r = 0.82$ (random selection of 14 of the 133 expert reports; Riegl, 2007) and $r = 0.84$ (random selection of 20 of the 339 expert reports; Adams, 2009), a correlation

coefficient of $r = 0.78$ (random selection of 51 of the 502 expert reports) was found in the present study.

2.3 Results

Table 2.1 presents the descriptive statistics for the total and sub scores. The mean total score was $M = 327.47$ ($SD = 62.62$), ranging from 202 to 435.5 points.

Table 2.1

Descriptive Statistics for the Total and Sub scores (N = 502)

Item	Score	Min	Max	M	SD
	Total score (Range 140 - 436.5)	**202**	**435.5**	**327.4**	**62.6**
	Sub score I (Range 12 - 41)	**18**	**41**	**34.55**	**4.83**
1.1	Order placement	1	3	2.98	0.17
1.2	Type of assessment report	1	3	2.97	0.19
1.4	Content presentation and overview	1	3	2.24	0.44
1.5	Information/description of the circumstances	1	3	2.29	0.47
1.6	Special examination and documentation methods	0	1	0.48	0.50
1.7	Accuracy and separate representation of the sources of knowledge	6	22	18.10	4.02
1.8	Linguistic representation	2	3	2.98	0.15
1.9	Documentation of the legal education of examinees	1	3	2.51	0.51
	Sub score II (Range 54 - 162)	**73**	**162.0**	**126.8**	**21.8**
2.1	Personal data	1	3	2.43	0.50
2.2	Offense	2	3	2.97	0.16
2.3	Profession & Finances	1	3	2.22	0.49
2.4	Medical histories/anamneses	4	12	10.73	1.95
2.5	Family of origin	7	21	14.35	4.26
2.6	Partnership and personal relationships	6	18	12.97	3.66
2.7	Sexual anamneses	4	12	7.74	3.24
2.8	Analysis of individual delinquency (backgrounds and causes)	14	42	33.32	8.80
2.9	Explored life spans	3	9	8.69	0.93
2.10	Sources of information used	5	15	11.37	2.49

Item	Score	Min	Max	M	SD
2.11	Identified relevant sources of information of the report	4	6	5.66	0.75
2.12	Care in handling the data/information validation	6	18	14.43	3.09
	Sub score III (Range 54 - 173.5)	**67**	**173.5**	**119.9**	**30.3**
3.1	Reference to further/other expert reports	2	6	5.64	0.78
3.2	Multidimensional data collection	3	9	6.25	2.37
3.3	Behavioral observation	5	15	11.88	2.60
3.4	Protective factors and risk factors	4	12	7.36	1.90
3.5	Delinquency genesis	4	12	7.58	2.34
3.6	Use of psychological testing and risk assessment instruments	1	4	2.50	1.19
3.7	Description of tests/instruments used and data obtained	1	12	4.99	4.14
3.8	Matching emp. knowledge about risk of recidivism of offender groups	20	62	41.24	17.3
3.9	Specified (current) conditional factors	5	15	10.56	2.46
3.10	Assessment of therapy/treatment progress since the index offense	4	12	9.89	2.12
3.11	Diagnostics	5	15	12.02	2.85
	Sub score IV (Range 20 - 60)	**20**	**60**	**46.15**	**9.77**
4.1	Frame of reference/limits of prognostic conclusion	5	15	10.08	2.82
4.2	Assessment of recidivism risk/loosening abuse	15	45	36.07	8.23

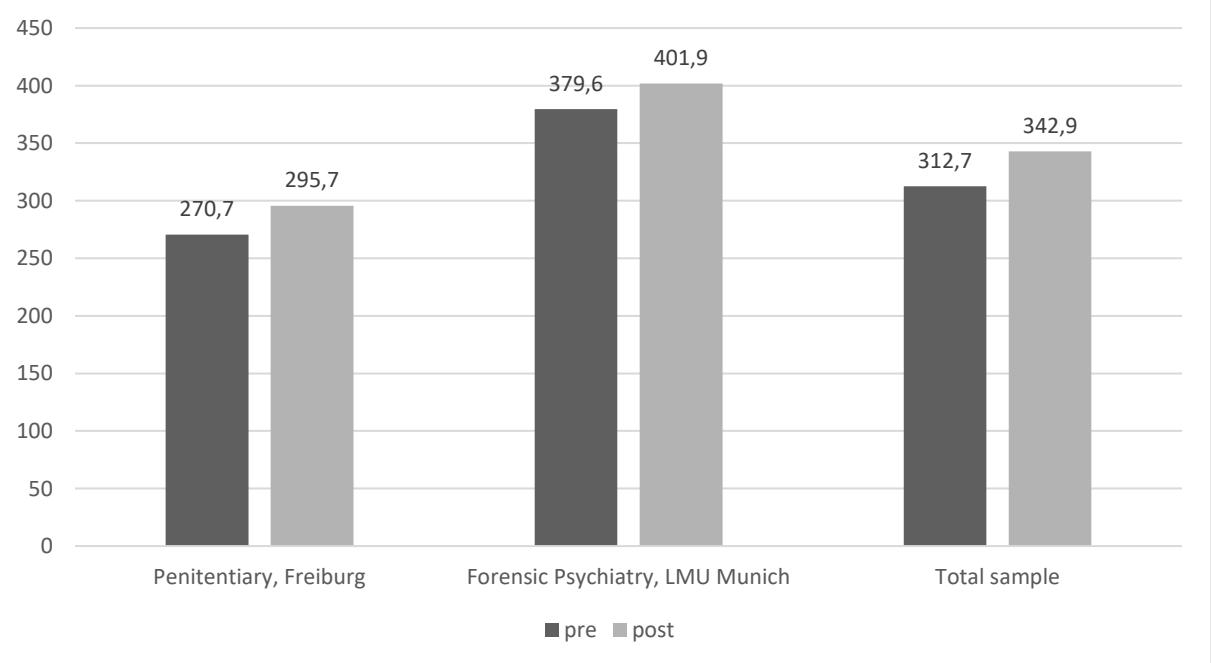
2.3.1 Application of the Methodological Minimum Requirements over time

To examine a potential improvement in the quality of risk assessment reports before and after the publication of the minimum requirements, reports that were conducted in the years 1999 to 2002, i.e. before publication (pre) were compared with those from 2008 to 2011 (post). The total sample of $N = 502$ was divided into $n = 257$ assessments before (pre) and $n = 245$ assessments after (post) the publication of the methodological minimum requirements. There was a significant mean difference regarding the total score between the assessment reports before ($M = 312.7$; $SD = 60.1$) and after ($M = 342.9$; $SD = 61.6$) the publication of the minimum requirements with respect to the total sample, $t(497.3) = -5.59$, $p < .001$. Such differences were also evident for the sub samples of the penitentiary in Freiburg and the

Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig-Maximilians University of Munich. Effect sizes of $d_{Cohen} = 0.5$ to $d_{Cohen} = 0.7$ were conducted for the total sample and the penitentiary in Freiburg, while the LMU Forensic Department showed an even larger effect of $d_{Cohen} = 1.1$. A more differentiated analysis of the total score showed a significant improvement in quality of reports of the total sample in all four sub scores *content and form*, $t(486.1) = -4.84$, $p < .001$, *procurement and handling of examinee-related information*, $t(497.2) = -5.40$, $p < .001$, *methodology*, $t(496.6) = -5.14$, $p < .001$, and *final conclusion*, $t(496.0) = -5.25$, $p < .001$. Effect sizes ranged from $d_{Cohen} = 0.43$ to $d_{Cohen} = 0.48$ (cf. Figure 1; Wertz & Kury, 2017a) since methodological minimum requirements were published.

Figure 2.1

Mean Comparisons of the Total Scores between Reports before and after the Publication of Methodological Minimum Requirements for the Total Sample and the Institutional Sub Samples (Kury & Wertz, 2017; Wertz & Kury, 2017a)



Comparing the present data with the results of the preliminary studies revealed no significant differences between the average quality of the risk assessment reports of the

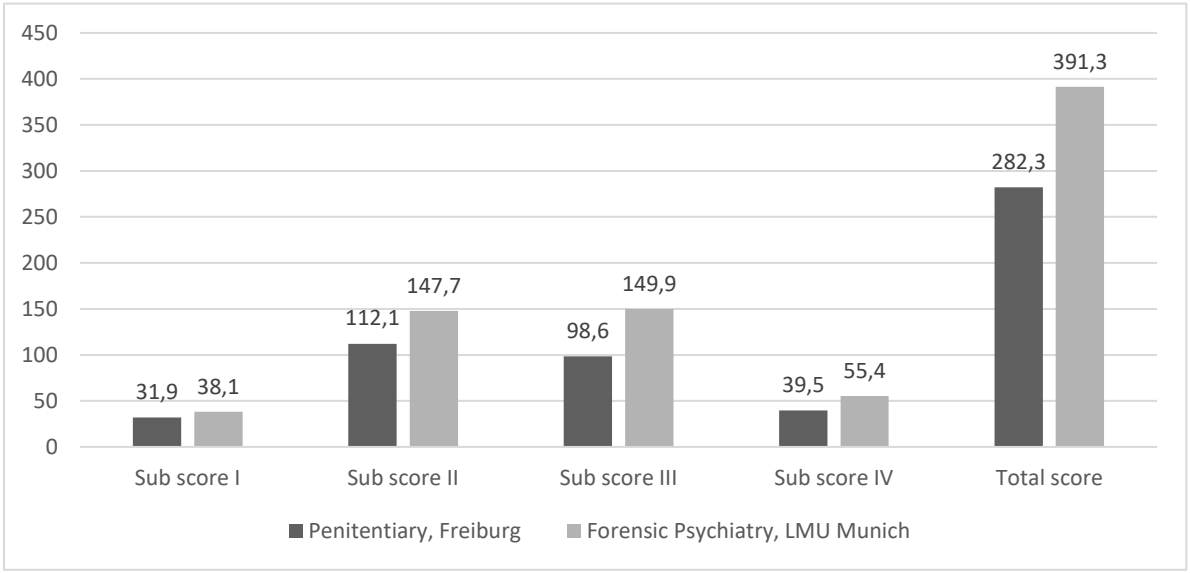
penitentiary in Freiburg conducted before 2006 (pre) and the results of Riegl (2007) and Adams (2009), $t(157) = 1.83$; $p = .069$ and $t(157) = -0.51$; $p = .612$. In contrast, the assessments provided after the publication in 2006 differed significantly from the results of the two authors, $t(135) = 6.92$, $p < .001$ and $t(157) = 8.70$, $p < .001$, respectively (Kury & Wertz, 2017; Wertz & Kury, 2017a, 2017b).

2.3.2 Application of the Methodological Minimum Requirements in Institutional Comparison

Within the institutional comparison, a significant mean difference within the total score between the assessment reports of the penitentiary in Freiburg (M = 282.3, SD = 37.5) and the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig Maximilian University of Munich (M = 391.3, SD = 22.4) could be shown, indicating a higher quality of reports of the university department, t(487.615) = -40.63, p < .001. Effect size measures showed a value of $d_{Cohen} = 3.3$, which can be classified as large (Wertz & Kury, 2017a).

Figure 2.2

Mean Comparisons of the Total and Sub Scores of Assessment Report Quality between the two included Institutions (Wertz & Kury, 2017a)



A more sophisticated analysis revealed significant higher assessment report quality in all four sub scores regarding *content and form*, $t(480.159) = -19.82$, $p < .001$, *procurement and handling of examinee-related information*, $t(497.548) = -32.36$, $p < .001$, *methodology*, $t(493.467) = -36.41$, $p < .001$, and *final conclusion*, $t(494.077) = -32.62$, $p < .001$, each in favour of the university department ($d_{Cohen} = 1.65$ to $d_{Cohen} = 3.06$).

Experts of the Department of Forensic Psychiatry showed a more pronounced homogeneity in quality. The average total score of all 17 examiners was $M = 396.4$ ($SD = 18.5$), which was significant higher than that of the 71 external experts of the penitentiary in Freiburg ($M = 275.8$, $SD = 18.2$). The average total scores of the university department ranged from 368.3 to 419.5 total points, forming a range of 51.2 scoring points. The external experts of the penitentiary in Freiburg revealed average total scores ranging from 202 to 389.5 points, forming a range of 187.5 points, which was almost five times higher compared to the score within the Forensic Psychiatry.

2.3.3 Application of the Methodological Minimum Requirements Depending on Profession, Prognostic Direction, Prognostic Issue, and Offense

Regarding differences within the quality of assessment reports depending on profession of assessors, a significant mean difference within the total scores between assessments conducted by psychiatrists ($n = 376$, $M = 337.3$, $SD = 65.2$) and psychologists ($n = 100$, $M = 307.4$, $SD = 42.1$) could be shown, resulting in an effect size of $d_{Cohen} = 0.49$, $t(240.058) = 5.53$, $p < .001$. There was also a significant difference between assessments within the penitentiary in Freiburg provided by psychiatrists ($n = 170$, $M = 271.9$, $SD = 32.2$) and psychologists ($n = 98$, $M = 305.1$, $SD = 39.3$), $t(171.774) = -7.09$, $p = .003$, showing an effect of $d_{Cohen} = 0.95$. Results provided evidence that reports conducted by psychologists were more closely aligned with the minimum requirements than those conducted by psychiatrists. A more differentiated analysis of the total score showed a significantly higher quality of assessments provided by psychologists in all four sub scores: *content and form*,

t(159.257) = -7.34; $p < .001$, $d_{Cohen} = 1.01$, *procurement and handling of examinee-related information*, t (179,350) = -5.56, $p < .001$; $d_{Cohen} = 0.73$, *methodology*, *t*(183,139) = -7.12, $p < .00$, $d_{Cohen} = 0.93$, and *final conclusion*, *t*(266) = -2.047, $p = .042$; $d_{Cohen} = 0.05$. Separate analysis of the assessment reports within the Department of Forensic Psychiatry at the Ludwig-Maximilians University of Munich was omitted due to the small group sizes (Kury & Wertz, 2017; Wertz & Kury, 2017b). Furthermore, institution-affiliated external experts within the penitentiary in Freiburg ($n = 175$, $M = 287.5$) showed significantly higher total scores than experts without institutional affiliation ($n = 119$, $M = 274.7$; Kury &Wertz, 2017).

While in the total sample no significant differences of quality ($N = 502$, *t*(449.858) = -0.14; $p = .887$, $d_{Cohen} = 0.01$; Kury & Wertz, 2017) could be found regarding the direction of prognostic conclusion (positively or negatively directed risk assessments), highly significant differences in favour of a higher quality for positively directed risk assessments were found in the subsample within the penitentiary in Freiburg, *t*(278,834) = 3.30, $p = .001$, $d_{Cohen} = 0.38$). In addition, a significant quality increase was found for assessment reports regarding a release of examinee compared to granting privileges, *t*(435) = 3.23, $p = .001$, $d_{Cohen} = 0.31$). Similarly, a significant mean difference was found with regard to the index delinquency, *F*(3, 498) = 4.78, $p = .003$, whereby the assessment reports about individuals convicted of sexual offenses showed the highest average total score and expert reports about individuals convicted of index offenses of murder and manslaughter performed the worst, $p = 0.003$.

2.3.4 Hit Rates of Positively Directed Risk Assessments According to Officially Registered Reoffenses

The hit rates of positively directed risk assessments according to the Federal Central Register were examined based on the profession of experts, over time and in relation to the quality of the assessment reports. Therefore, a positively directed risk assessment was defined as an affirmation of the examinee´s release or granting of privileges. By analogy, *accurate* was described as a positively directed risk assessment that was correct, and *inaccurate* was

described as a recidivism occurred according to the Federal Central Register despite a positively directed prognosis. Negatively directed assessment reports were not taken into account for the calculation of the hit rates, since a negatively directed risk assessment (assuming that the judicial decision generally implements the expert's recommendation) was usually not associated with any detention measures or releases and therefore no possibility of recidivism was detected. A distinction was made between general recidivism (at least one entry in the Federal Central Register since the assessment, irrespective of the type and severity of the recidivism), serious recidivism (conviction to a new custodial sentence) and relevant recidivism (an offense directed against the same legal interest as the index offense). Group differences were found for the total sample depending on the profession of the experts for general, $\chi^2(1, 243) = 12.37$, $p < .05$, $d_{Cohen} = .46$, and severe recidivism, $\chi^2(1, 243) = 12.27$, $p < .05$, $d_{Cohen} = .46$), in favour of the psychiatrists. The hit rate for relevant relapse showed the same trend but without statistical significance, $\chi^2(1, 243) = 1.54$, $p > .05$, $d_{Cohen} = .16$.

Within the sub sample of the external reports of the penitentiary in Freiburg, there were no significant differences regarding the hit rates between psychiatric and psychological assessments, although it was found to be a significantly more restrictive approach by the psychiatrists, who provided significantly less positively directed risk assessments overall. Psychiatric experts provided a negatively directed risk assessment in 92 of 170 evaluations (54.1 %.), because the inmate was still considered dangerous enough that release or granting of privileges could not be recommended. For psychologists, on the other hand, this proportion was significantly lower, resulting in 20 out of 98 examinations (20.4%). Furthermore, significant differences in hit rates between the two institutions could be found for general, $\chi^2(1, 258) = 21.88$, $p < .05$, $d_{Cohen} = .61$, and severe recidivism, $\chi^2(1, 258) = 12.90$, $p < .05$, $d_{Cohen} = .46$, in favour of the university department; the hit rates for relevant

recidivism showed no statistically significant differences (cf. Table 2.2; Wertz & Kury, 2017b).

Table 2.2

Results of χ^2-tests and Frequencies of (in)accurate positively directed Risk Assessments depending on Institutions for general, severe, and relevant Recidivism (N = 258; Wertz & Kury, 2017b)

Variable	General recidivism[a]		Severe recidivism[b]		Relevant recidivism	
	Freiburg	Munich	Freiburg	Munich	Freiburg	Munich
Accurately positive	77	79	112	91	145	94
Inaccurately positive	80	22	45	10	12	7
Hit rate (%)	49.0	78.2	71.3	90.1	92.4	93.1
n	157	101	157	101	157	101

Note.

[a] $\chi^2 = 21.88$, $df = 1$, $p < .05$, $d = .61$

[b] $\chi^2 = 12.90$, $df = 1$, $p < .05$, $d = .46$

Overall, there was also a significant positive correlation between the total score of application of the methodological minimum requirements for risk assessment reports and the hit rates of positively directed risk assessments, $r = .25$, $p < .01$, (Kury & Wertz, 2017).

2.4 Discussion

2.4.1 Quality Improvement Since the Publication of Methodological Minimum Requirements

The comparison of the risk assessment reports before and after the publication of the methodological minimum requirements (Boetticher et al., 2006) shows that there has been a significant improvement in quality over time for the overall sample and for the included institutions separately. Thus, the developed methodological minimum requirements seem to have arrived in practice, but there is still an enormous potential for improvement concerning

the practical application of the criteria. The results are in line with the studies by Herschbach (2009), Passow (2010) and Schläfke et al. (2000), who also analysed risk assessment reports within certain cohorts and, despite considerable quality deficiencies, generally found an improvement in the quality within the reports (Wertz & Kury, 2017a).

2.4.2 Special Status of University Institutions

With regard to the differences concerning the institutional comparison in favour of the university hospital, the involvement of the department management in the publication of the minimum requirements as well as the research proximity of a university department can be discussed, since university institutions have a special status in the review practice. This conclusion is also supported by the observed differences in quality depending on the institutional affiliation of the external reviewers at the penitentiary in Freiburg. Accordingly, in the context of institutional quality assurance, an exchange of expertise among colleagues, access to current research literature, and mandatory training and continuing certification programs can be guaranteed (Wertz &Kury, 2017a).

2.4.3 Research Proximity More Decisive Than Profession

Results showed a significant higher quality of reports provided by psychologists at the penitentiary in Freiburg than reports conducted by psychiatrists. Reports conducted by psychologists were more closely aligned with the minimum requirements regarding the total score and the four subscores. However, taking into account the entire sample of both institutions, a significant difference pointing to a higher quality of expert reports prepared by psychiatrists was found. Considering the operationalization of expert witness report quality via compliance with the minimum requirements, it must be discussed whether educational and further trainings as well as the institutional affiliation of the experts may be not of greater relevance than the corresponding profession of the experts. This assumption is also suggested by the high quality of assessment reports of both professions. Overall, results indicate that both professions, psychiatrists, and psychologists, should have at least an equal position as

experts, with results showing that psychiatrists are primarily used as experts and psychologists are often consulted for psychological test evaluations only - despite their corresponding (especially methodological and clinical-psychological) competencies.

2.4.4 Hit Rates of Positively Directed Risk Assessments According to Officially Registered Reoffenses

Finally, the quality of a risk assessment can also be analysed by whether the assessment can be considered accurate with regard to the future legal behavior of examinees. In the present study, only positively directed risk assessment reports could be analysed regarding hit rates. Accordingly, the proportion of false-positive assessments could not be examined in the present study, because there were no possibilities for the individuals classified as still too dangerous to prove their potential non-dangerousness outside the prison or psychiatric hospital system. Studies in which such questions have been addressed (cf. e.g., Alex, 2010; Müller & Stolpmann, 2015) consistently showed a relatively high proportion of false-positive assessments. Validation of the hit rates showed significant group differences for general and severe recidivism in reports provided by psychiatrists compared to psychologists within the total sample, with only a (non-significant) trend for relevant recidivism. Accordingly, results support the presented differences regarding the quality between the two professions. These differences must be considered in combination with the different orientations of the two institutions being compared. Results were also confirmed by the group differences regarding the hit rate for general and severe recidivism in favour of the university department of forensic psychiatry compared to the penitentiary in Freiburg. Within the subsample of the penitentiary, however, there were no significant differences regarding the hit rates between external psychiatric and psychological assessments, which should be emphasized, especially considering the much more restrictive approach of the psychiatrists. In 92 of 170 expert reports (54.1%), the psychiatric experts at the correctional facility came to a

negatively directed risk assessment, while this proportion for the psychologists was significantly lower with 20 out of 98 negatively directed risk assessment (20.4%).

2.4.5 Relation between Quality and Hit Rates

The occurrence of inaccurate risk assessments can be reduced, but not eliminated by the compliance with the minimum requirements for risk assessment reports (Dahle, 2005b; 2006). Empirical evidence for a correlation between prognostic failures and prognostic mistakes could not be found in the current research literature so far. To examine whether expert reports, which are more closely oriented to the methodological minimum requirements, also show higher hit rates, expert reports were compared to the results of the quality ratings of the assessment reports. It was found that compliance with the methodological minimum requirements is associated with a higher accuracy of the risk assessments (Wertz & Kury, 2017a; 2017b). It can be concluded that experts should be encouraged more consistently to consider and apply the minimum requirements, as these clearly contribute to an increase in the quality of the risk assessment reports and thus to a more valid prognostic assessment of the recidivism risk of the examinees.

2.4.6 Limitations

The methodological limitations of the non-validated survey instrument due to the lack of psychometric quality and the restrictive three-level scaling of the questionnaire are offset by numerous possibilities regarding a precise differentiation that allows an operationalization of the quality of the risk assessment reports as applied in this study. Reliability, on the other hand, was measured by a second assessment of a co-rater and can be classified as satisfying. Reference has already been made to the comparability of the content of the two institutions emphasising the special status of institutions such as universities. Studies on the validity of risk assessment reports are likewise subject to limited interpretation, since the legal consequences of a negatively directed prognosis often lead to lengthy custodial measures, making it impossible to prove false-positive assessments (Dittmann, 2012). Moreover, Federal

Central Register excerpts represent only one of several possible recidivism data sources, where a systematic underestimation of actual recidivism can usually be assumed.

3. The Influence of Judicial Orders on the Quality of Criminal Risk Assessment Reports[3]

Abstract

According to the minimum requirements and current recommendations regarding criminal risk assessments, judicial orders are required to explicitly formulate the expert's assignment, to describe the assessment's matter, and to clarify the exact questions the expert is supposed to answer. The assessment should be oriented along four fundamental prognostic questions: (1) the probability of re-offenses, (2) the nature, frequency, and severity of such offenses, (3) possible risk-reducing interventions, and (4) possible risk-increasing circumstances. There is no empirical evidence as to which extent these requirements are implemented by judicial orders, nor if or how expert statements answer these fundamental questions in their criminal risk assessments. In this study, a retrospective analysis of risk assessment reports of individuals convicted of sexual and/or violent offenses from the penitentiary in Freiburg and the Department of Forensic Psychiatry of the University Hospital Munich (N = 787) between the years of 1999 until 2006 was conducted. These assessment reports underwent an empirical validation regarding the answering of these fundamental prognostic questions in relation to the judicial order that was formulated. The results showed that since the methodological minimum requirements were published, the judicial orders incrementally followed the standards regarding the formulation of the assessment assignment, and as a result expert's increasingly answered the fundamental prognostic questions that were posed. Still, it was visible that clinical and judicial practice were rather heterogeneous. Overall, a statistically significant relationship between the manner in which judicial orders were formulated and the quality of the corresponding expert statement could be established.

[3] Paper published as Wertz, M., Schiltz, K., Imhoff, R.& Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosebegutachtung [The influence of the judicial order on the quality of the work of expert witnesses in the context of risk assessment]. Recht & Psychiatrie, 38(4), 193 – 200.

Specifically, accurate phrasing of the questions of interest led to an extensive answer in the expert reports.

Keywords: criminal risk assessment, minimum requirements, prognostic questions, judicial order, expert witness reports

3.1 Answering of Judicial Orders in Criminal Risk Assessments: (Legal) Framework and Empirical Recommendations

In the relevant legal texts on the judicial system, various and sometimes vague formulations of expert´s assignments regarding criminal risk assessment can be found (Dahle, 2010; Boetticher et al. 2019). However, in order to be able to meet the expectations and needs of the recipients, judicial orders are required to explicitly formulate the expert´s assignment in order to clarify which questions should and should not be addressed within risk assessment reports. Simplified, it can be concluded that "the more detailed the formulated assignment, the better the quality of the assessment report" (Böhm, 2018, p. 134).

In order to ensure more consistent answers to fundamental prognostic questions by experts despite these different legal requirements and matters of prognostic questions, specific methodological recommendations for risk assessment reports were published (Boetticher et al., 2019). Besides updating the methodological minimum requirements by an interdisciplinary working group (Boetticher et al., 2006), concrete fundamental questions to which "even the judicial order should at least be oriented " (p. 555) were formulated from a legal perspective:

1. What is the probability of re-offenses?
2. What will be the nature, frequency, and severity of these crimes?
3. What possible risk-reducing interventions can be taken?
4. What are possible risk-increasing circumstances?

The judicial order should "precisely describe the assessments matter and clarify which questions are actually to be answered by the expert; the mere reproduction of the legal text is regularly not sufficient for this purpose" (Boetticher et al., 2006, p. 539).

There is no empirical evidence as to which extent these requirements are implemented by judicial orders, nor if or how expert statements answer these fundamental questions in their criminal risk assessments. There is also no data available to date whether the publication of the methodological minimum requirements for risk assessment reports in 2006 has led to greater compliance with these requirements by orders and experts. Furthermore, there is no evidence of the influence of judicial orders on the quality of risk assessment reports. As the recommendations for risk assessment reports are also intended to contribute to a dialogue capability between orderers and experts (Nedopil, 2005; Pfister, 2019), it seems relevant to examine to what extent explicit formulations of judicial orders influence the quality of risk assessment reports in general practice.

3.2 Methods

In order to examine the extent to which these requirements are implemented by judicial orders (Boetticher et al., 2006; 2019) and how expert statements answer these fundamental questions in their criminal risk assessments, risk assessment reports about individuals convicted of sexual and violent offenses (N = 787) underwent an empirical investigation. By analysing the concrete formulations of judicial orders, the compliance with the methodological minimum requirements was examined. Results are presented regarding the year of contribution of reports, the accommodation or imprisonment situation of individuals, the index offenses, the diagnoses, the profession of the experts, the interdisciplinarity of reports (in the form of the involvement of additional psychological test reports or social scientific explorations), and the institutions included, as well as over time - before or after the publication of the first edition of the methodological minimum requirements in 2006.

3.2.1 Sample description and descriptive statistics

The analysis units of the present study were *n* = 412 external risk assessment reports about inmates from the penitentiary in Freiburg and *n* = 375 reports about individuals convicted of sexual and violent offenses from the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of Ludwig-Maximilian University in Munich conducted between 1999 and 2016 (*N* = 787).

Of the risk assessment reports provided by the university department in Munich, *n* = 154 (41.1%) were conducted before the publication of the methodological minimum requirements, and *n* = 221 (58.9%) were conducted after (Boetticher et al. 2006). Of the external reports provided by the penitentiary in Freiburg, *n* = 253 (61.4%) were conducted before, and *n* = 159 (38.6%) after 2006. The total number of assessment reports were ordered by diverse judicial parties to the proceedings (e.g. local or district courts, or higher regional courts). In the university department, a total of *n* = 25 different department-affiliated experts were involved in the conduction of assessment reports, whereas in the penitentiary in Freiburg, a total of *n* = 85 external experts from independent expert practices, forensic psychiatric hospitals, psychological institutes, or psychotherapeutic and psychiatric hospitals or care centres were consulted. While at the Department of Forensic Psychiatry reports were conducted almost exclusively by psychiatric experts (*n* = 24, 96.0%), in the penitentiary in Freiburg external psychologists (*n* = 17, 20.0%), psychiatrists (n = 58, 68.2%), and experts of both professions (*n* = 10, 11.8%) were asked to assess examinees. Thus, of the total *N* = 787 assessment reports analysed, *n* = 620 were provided by medical experts, *n* = 137 by psychological experts, and *n* = 30 by experts of both professions.

Of the total sample of *n* = 375 expert reports from the university department, *n* = 347 (92.5%) assessments were provided interdisciplinarily (by consulting an additional psychological test examination or an additional exploration by a sociologist). In under 10% (*n* = 40) of the expert reports of the penitentiary in Freiburg an interdisciplinary procedure was

documented. The statistical sample description presents the judicial orders, the accommodation/imprisonment situation of examinees, the index offenses, and the psychiatric diagnoses according to ICD-10 depending on the included institution and time period of contribution (cf. Table 3.1).

Table 3.1

Statistical Sample Description depending on Institution and Time Period of Contribution

	LMU, Munich[a] (*n* = 375)				Penitentiary, Freiburg (*n* = 412)			
	pre[b] (*n* = 154)		post[b] (*n* = 221)		pre[b] (*n* = 253)		post[b] (*n* = 159)	
	n	*%*	*N*	*%*	*n*	*%*	*n*	*%*
Accommodation/Imprisonment								
Penitentiary	54	35.0	44	20.0	188	74.0	114	72.0
Preventive Detention § 66	25	16.0	34	15.0	65	26.0	45	28.0
Forensic Psychiatry § 63	69	45.0	137	62.0	-	-	-	-
Forensic Psychiatry § 64	6	4.0	5	2.0	-	-	-	-
None	-	-	1	<1.0	-	-	-	-
Judicial Order								
Granting Privileges	10	6.0	5	2.0	76	30.0	61	38.0
Release	75	49.0	99	45.0	98	39.0	43	28.0
Privileges/Release	42	27.0	76	34.0	3	1.0	-	-
Preconditions for §66	5	3.0	12	5.0	41	16.0	34	21.0
Privileges of §66	6	4.0	1	<1.0	19	8.0	13	8.0
Release of §66	9	6.0	18	8.0	15	7.0	8	5.0
Privileges/Release of § 66	7	5.0	10	5.0	1	<1.0	-	-
Index Offence								
Violent Offence	61	40.0	104	47.0	158	62.0	98	61.0
Sexual Offence	22	14.0	43	19.0	58	23.0	32	20.0
Violent-/Sexual Offence	54	35.0	50	22.0	28	11.0	27	16.0
Property-/Fraud Offence	8	5.0	6	3.0	7	3.0	4	2.0
Offence by narcotics law	2	1.0	8	4.0	2	1.0	2	1.0
Arson	7	5.0	10	5.0	-	-	-	-
Psychiatric Diagnoses (ICD-10)								
No Diagnoses	55	36.0	45	20.0	145	57.0	107	67.0
F00-F09	4	3.0	9	4.0	-	-	1	1.0
F10-F19	8	5.0	14	6.0	25	10.0	9	6.0
F20-F29	20	13.0	40	18.0	-	-	-	-
F30-F39	-	-	8	4.0	-	-	-	-
F40-F49	-	-	2	1.0	-	-	1	1.0
F50-F59	-	-	-	-	-	-	-	-
F60-F69	62	40.0	85	38.0	83	33.0	41	25.0
F70-F79	5	3.0	15	7.0	-	-	-	-
F80-F89	-	-	1	<1.0	-	-	-	-
F90-F99	-	-	2	1.0	-	-	-	-

Notes. *N* = 787.

[a] LMU Munich (Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy at the Ludwig-Maximilians-University Munich)

[b] pre/post (before and after publication of methodological minimum requirements in 2006, Boetticher et al. 2006)

3.2.2 Empirical validation of the accordance between judicial order and answering in expert witness reports

To empirically examine the research question, judicial orders and the answers to the prognostic questions in the risk assessment reports were systematically gathered. It was examined whether a stronger orientation to the methodological minimum requirements can be found over time and to what extent there is a relation between judicial orders and their answers in risk assessment reports.

3.2.3 Examination of judicial orders

The presence of each of the four fundamental prognostic questions in judicial orders was gathered dichotomously ("yes" = consideration of the question in the judicial order, "no" = no consideration in the judicial order). An overview was then drawn up for each judicial order as to whether it referred separately to (1) the probability of re-offenses, (2) the nature, frequency, and severity of such offenses, (3) possible risk-reducing interventions as well as (4) possible risk-increasing circumstances. In all assessment reports, the exact questions of the judicial order were reported on the first page with reference to the orderer or indication of the judicial decisions. It could thereby be reliably assumed that the corresponding judicial orders were included literally in the introductory part of the reports, as also required by the methodological minimum requirements or recommendations for risk assessment reports (Boetticher et al. 2006; Kröber et al. 2019). In addition to the dichotomous evaluation of the consideration of the fundamental questions, the exact wording of the judicial orders was taken literally from the digitally archived reports of the university department in Munich (n = 375).

3.2.4 Examination of answers to fundamental prognostic questions in risk assessment reports

The expert's response to the four fundamental prognostic questions was rated on a three-level scale ("absent" = no reference to the corresponding question, "incomplete" = only partial reference to the question, "comprehensive" = complete reference to the question). The first prognostic question regarding the probability of re-offenses was rated as "incomplete" for a purely descriptive, unstructured description of risk with either no or subjective assessment, whereas a categorical (using risk categories) or quantitative (numerical) statement of probability was rated as "comprehensive response." In assessing the second prognostic question, the expert's response was rated as "absent" if no reference to type, frequency, or severity was found in the prognostic assessment, as "incomplete" if only one or two of these three aspects were addressed, and as "comprehensive" if all aspects of the entire question were addressed. For the third and fourth prognostic questions, naming at least one risk-reducing measure or one risk-increasing circumstance was sufficient for a rating of "incomplete." A response rated as "comprehensive" was assigned if more than one measure or circumstance was discussed. If an item could not be rated by the expert in the report due to comprehensible explanations and was documented accordingly, it was still rated as "comprehensive". To determine interrater reliability, a random sample of 10% (n = 78) of the total number of expert reports (Leonhart, 2004) was analysed independently of the first rater by an in-depth and intensively trained second rater. The analyses yielded reliability coefficients that can be classified as very high according to the usual standards (Leonhart, 2004), which means that interrater reliability can be regarded as fully given for all four fundamental prognostic questions: For meeting the probability criteria, it was *ICC* = .814, p < .005, 95% CI [.723-.839]; for nature, frequency, and severity, ICC = .916, p < .005, 95% CI [.871-.931], for risk-reducing measures it was *ICC* = .920, p < .005, 95% CI [.877-.935], and for risk-increasing circumstances it was *ICC* = .952, p < .005, 95% CI [.925-.961].

3.3 Results

First, the dichotomous evaluations of the fundamental prognostic questions are presented descriptively. Furthermore, frequently used word-for-word formulations of judicial orders were gathered out of the assessment reports as examples for illustration. In addition, the experts' answers to the four fundamental prognostic questions are presented depending on included institution and over time. Finally, answers are correlated with the previously mentioned judicial order variables.

3.3.1 Formulation of the judicial order

To answer the question of the extent to which judicial orders are guided by the four fundamental prognostic questions, those assessment reports conducted before the publication (pre; n = 407) of the methodological minimum requirements were compared with reports provided from 2007 to 2016 (post; n = 380). Significant pre-post differences could be found with respect to the frequency of consideration of (1) the probability of re-offenses, (2) the nature, frequency, and severity, and (4) risk-increasing circumstances, in favour of a more common consideration of these formulations in risk assessment reports provided after 2006 (cf. Table 3.2 for more details). The effect sizes calculated varied from low values (Cramérs φ = .11 and .15, respectively) for (1) the probability of re-offenses and (2) the nature, frequency, and severity to a very strong difference value of Cramérs φ = for .87 for (4) risk-increasing circumstances. No statistical differences over time could be found for the question of (3) risk-reducing measures.

Table 3.2

Consideration of the Four Fundamental Prognostic Questions in Judicial Orders before (pre) and after (post) the Publication of Methodological Minimum Requirements for Risk Assessment Reports in 2006

Answering to…	(1) Probability of re-offenses		(2) Nature, frequency, severity		(3) Risk-reducing measures		(4) Risk-increasing circumstances	
	no[b]	yes[b]	no[b]	yes[b]	no[b]	yes[b]	no[b]	yes[b]
pre[a]	107	300	378	29	203	204	362	45
post[a]	66	314	315	65	181	199	315	65
Statistics	$\chi^2(1) = 9.12$ $p = .003$ $\varphi = .11$		$\chi^2(1) = 18.61$ $p < .001$ $\varphi = .15$		$\chi^2(1) = .40$ $p = .529$ $\varphi = .02$		$\chi^2(1) = 5.98$ $p = .014$ $\varphi = .87$	

Notes. $N = 787$ (Probability of re-offenses: $n_{yes} = 614$, $n_{no} = 143$; Nature, frequency, severity: $n_{yes} = 94$, $n_{no} = 693$; Risk-reducing measures: $n_{yes} = 403$, $n_{no} = 384$; Risk-increasing circumstances: $n_{yes} = 110$, $n_{no} = 677$).

[a] pre/post (before and after publication of methodological minimum requirements, cf. Boetticher et al., 2006)

[b] no/yes ("no" = no consideration / "yes" = consideration in the judicial order)

In addition to the dichotomous evaluations, frequently used word-for-word formulations of judicial orders were gathered from the digitally archived risk assessment reports of the university department in Munich ($n = 375$). Irrespective of the necessary orientation to different legal contexts and specifications, identified wordings illustrate the heterogeneity of judicial orders for risk assessment reports. Thus, the judicial orders asked for "criminal", "social", or "legal risk", or even "dangerousness" of examinees or assessments of the "risk of re-offense" and therefore an associated "qualified" or "high risk of re-offenses", a "danger to the general public", the "expectation of new offenses", the "responsibility of granting privileges or of a release", the "personality in relation to dangerousness", the "maturity for release", the "existence of dangerousness to others", the "tendency to commit severe offenses", or "fears of escape and abuse in case of privileges or release". The expert

was expected to provide different degrees of assessment, i.e., whether re-offenses are "to be expected", whether they can be "excluded", or whether the risk can be "assumed".

Despite the significantly stronger orientation to the methodological minimum requirements over time presented above, the heterogeneity - also within the framework of the same legal standards - was reflected not only in the word-for-word formulations but also in the comprehensiveness of the orders. The total number of text characters in the judicial orders varied considerably between the various orderers, as illustrated by the following two examples: the first order with 80 text characters and the second one with 3.012 text characters. Both judicial orders refer equally to an order for granting privileges or release from psychiatric hospital according to § 63 of the German penal code:

- First example of a judicial order (from 2012): "An external expert witness report regarding diagnostics and prognosis is to be obtained"
- Abbreviated excerpt of a second example of a judicial order (from 2009): “According to the legal decision (...), the expert report should consider the following questions: Does the convicted person, from today's point of view, have a mental disorder in the sense of the requirements of § 20 of the German penal code? (...) What consequences does this disorder have on the risk assessment of the convicted person (§ 63 of the German penal code)? (...) If applicable, have the mental disorders of the convicted person and/or his tendency to excessive consumption of addictive substances undergone changes in the course of the treatment? If so, a) what are the nature of the changes and how do they manifest themselves? b) How do the changes affect the risk assessment of the convicted person c) Which risk factors persist? Besides, is there a need for further treatment, and what prospects does further treatment offer to further reduce the remaining risk factors and the risk of re-offenses of the convicted person. In particular, what promising short-, medium- and long-term treatment options are available in this respect, and which of these options appear preferable in the specific case? (...) Under what

> conditions would further treatment be possible and medically justifiable outside of inpatient placement in a forensic psychiatry? In what other way could the identified risk factors be mitigated in their significance or even completely compensated? How should the criminal risk of the convicted person be assessed from a forensic psychiatric perspective? In particular, can it be assumed today that there is no longer any risk that the convicted person will continue to relapse, and can it be expected that the convicted person will no longer commit any unlawful offenses outside of the forensic psychiatry? Can the probability that the convicted person will not re-offend outside the forensic psychiatry be secured or even further increased by accompanying and supporting measures (directives), and if so, by which and to what extent? (...) With what degree of probability and with what temporal range can the risk assessment be made?".

3.3.2 Answers to fundamental prognostic questions in expert reports

First, risk assessment reports conducted before and after the publication of the methodological minimum requirements were compared. For the total sample, significant group differences were found in answers to the prognostic questions (1) and (2) between expert reports conducted before and after the publication of the minimum requirements with effect sizes of Cramérs φ= .13 and .18, classified as low, whereas no statistically significant differences could be determined in the answers to questions (3) and (4) (cf. Table 3.3).

Table 3.3

Answers to the Four Fundamental Prognostic Questions before and after the Publication of Methodological Minimum Requirements for Risk Assessment Reports in 2006

Answering to...	(1) Probability of re-offenses		(2) Nature, frequency, severity		(3) Risk-reducing measures		(4) Risk-increasing circumstances	
	pre[a]	post[a]	pre[a]	post[a]	pre[a]	post[a]	pre[a]	post[a]
absent[b]	38	32	200	145	48	40	93	92
incomplete[b]	162	91	172	180	87	91	124	112
comprehensive[b]	207	257	35	55	272	249	190	176
Statistics	$\chi^2(2) = 24.93$ $p < .001$ $\varphi = .18$		$\chi^2(2) = 12.48$ $p = .002$ $\varphi = .13$		$\chi^2(2) = .91$ $p = .635$ $\varphi = .04$		$\chi^2(2) = .23$ $p = .894$ $\varphi = .02$	

Notes. $N = 787$ (Probability of re-offenses: $n_{pre} = 407$, $n_{post} = 380$; Nature, frequency, severity: $n_{pre} = 407$, $n_{post} = 380$; Risk-reducing measures: $n_{pre} = 407$, $n_{post} = 380$; Risk-increasing circumstances: $n_{pre} = 407$, $n_{post} = 380$).

[a] pre/post (before and after publication of methodological minimum requirements, cf. Boetticher et al., 2006)

[b] absent (no reference to the question), incomplete (only partial reference to the question), comprehensive (complete reference to the question)

While the sub-sample of the risk assessment reports of the university department showed a statistically significant increase in the frequency and comprehensiveness of responses to questions (1), (2), and (4) after 2006, with effect sizes that were considered small to moderate, Cramérs φ = .22 to .31 (cf. Table 3.4), the external risk assessment reports of the penitentiary in Freiburg did not show a significantly more frequent or comprehensive response to any of the four fundamental prognostic questions over time.

Table 3.4

Answering to the Four Fundamental Prognostic Questions in Risk Assessment Reports of the University Department in Munich before and after the Publication of Methodological Minimum Requirements for Risk Assessment Reports in 2006

Answering to...	(1) Probability of re-offenses		(2) Nature, frequency, severity		(3) Risk-reducing measures		(4) Risk-increasing circumstances	
	pre[a]	post[a]	pre[a]	post[a]	pre[a]	post[a]	pre[a]	post[a]
absent[b]	4	13	79	60	18	12	24	10
incomplete[b]	72	41	70	118	32	41	61	78
comprehensive[b]	78	167	5	43	104	168	69	133
Statistics	$\chi^2(2) = 34.74$ $p < .001$ $\varphi = .31$		$\chi^2(2) = 34.05$ $p < .001$ $\varphi = .31$		$\chi^2(2) = 5.58$ $p = .062$ $\varphi = .12$		$\chi^2(2) = 16.68$ $p < .001$ $\varphi = .22$	

Notes. $N = 375$ (Probability of re-offenses: $n_{pre} = 154$, $n_{post} = 221$; Nature, frequency, severity: $n_{pre} = 154$, $n_{post} = 221$; Risk-reducing measures: $n_{pre} = 154$, $n_{post} = 221$; Risk-increasing circumstances: $n_{pre} = 154$, $n_{post} = 221$).

[a] pre/post (before and after publication of methodological minimum requirements, cf. Boetticher et al., 2006)

[b] absent (no reference to the question), incomplete (only partial reference to the question), comprehensive (complete reference to the question)

Comparing the included institutions, significant group differences in the answering of all four fundamental prognostic questions between risk assessment reports of the two institutions were found, showing more frequent and comprehensive responses in reports of the university department with effect sizes of Cramérs φ= .13 to .33 characterized as low to moderate, cf. Table 3.5.

Table 3.5

Answering to the Four Fundamental Prognostic Questions in Risk Assessment Reports depending on the included Institutions

Answering to...	(1) Probability of re-offenses		(2) Nature, frequency, severity		(3) Risk-reducing measures		(4) Risk-increasing circumstances	
	PEN[a]	DFP[a]	PEN[a]	DFP[a]	PEN[a]	DFP[a]	PEN[a]	DFP[a]
absent[b]	53	17	206	139	58	30	151	34
incomplete[b]	140	113	164	188	105	73	97	139
comprehensive[b]	219	245	42	48	249	272	164	202
Statstics	$\chi^2(2) = 21.16$ $p < .001$ $\varphi = .17$		$\chi^2(2) = 13.34$ $p < .001$ $\varphi = .13$		$\chi^2(2) = 13.97$ $p < .001$ $\varphi = .14$		$\chi^2(2) = 83.86$ $p < .001$ $\varphi = .33$	

Notes. $N = 787$; (Probability of re-offenses: $n_{PEN} = 412$, $n_{DFP} = 375$; Nature, frequency, severity: $n_{PEN} = 412$, $n_{DFP} = 375$; Risk-reducing measures: $n_{PEN} = 412$, $n_{DFP} = 375$; Risk-increasing circumstances: $n_{PEN} = 412$, $n_{DFP} = 375$).

[a] PEN (Penitentiary, Freiburg), DFP (Department of Forensic Psychiatry, Munich)

[b] absent (no reference to the question), incomplete (only partial reference to the question), comprehensive (complete reference to the question)

Analyses of variance revealed significant main effects of institution, $p < .001$, with respect to responses to all four prognostic questions (1 - 4) and time period (pre/post) on responses to (1) the probability of re-offenses, $p = .001$, and (2) the nature, frequency, and severity, $p = .002$, of reoffending. To test for specific interactions in addition to the two main effects, additional two-factor ANOVAs were calculated for all four dependent variables. While responses to (1) the probability and (2) the nature, frequency, and severity of re-offenses showed only additive main effects but no interaction, $F(1,783) = 2.40$, $p = .122$, and $F(1,783) = 2.85$, $p = .092$, both the (3) risk-reducing measures, $F(1,783) = 13.303$, $p < .000$, partial $\eta 2 = .035$, and the (4) risk-increasing circumstances, $F(1,783) = 41.940$, $p < .000$, partial $\eta 2 = .051$, showed significant interactions of the two factors. Consequently, looking in

detail at the differences presented due to institution and time period (cf. Tables 3.3, 3.4, and 3.5), the effect between institutions seemed to have increased after the publication of the methodological minimum requirements, i.e., the recommendations may have been taken up more in the research-related university institution than in the general clinical assessment practice.

3.3.3 Relation between judicial orders and the answering of the corresponding prognostic questions in risk assessment reports

To examine whether the degree of response to the relevant prognostic questions by experts depended on the explicit questions of the judicial order, biserial rank correlations were calculated both for the total sample and separately by institution (cf. Table 6). Except for a statistically non-significant rank correlation for the sub-sample of the university department with regard to ordering and answering of question (1), the correlations were consistently in line with expectations, suggesting that the formulation of the judicial orders led to a stronger orientation towards these questions in corresponding risk assessment reports.

Table 3.6

Biserial Rank Correlations of the asked Questions in Judicial Orders and their Answers by Experts regarding the Four Fundamental Prognostic Questions depending on the Institution

Asking for...	(1) Probability of re-offenses		(2) Nature, frequency, severity		(3) Risk-reducing measures		(4) Risk-increasing circumstances	
	r_{PEN}[a]	r_{DFP}[a]	r_{PEN}[a]	r_{DFP}[a]	r_{PEN}[a]	r_{DFP}[a]	r_{PEN}[a]	r_{DFP}[a]
Answering to...								
(1)	.30 **	.05	.25 **	.22 **	.00	.09	.14 **	.12 **
(2)	.13 **	.10*	.38*	.33 **	.11*	.15 **	.23**	.27*
(3)	.05	.05	.00	.11*	.43 **	.21 **	.16 **	.15*
(4)	.05	.08	.12*	.15*	.12*	.17**	.42**	.22*

Notes. N = 787 (Probability of re-offenses: n_{PEN} = 412, n_{DFP} = 375; Nature, frequency, severity: n_{PEN} = 412, n_{DFP} = 375; Risk-reducing measures: n_{PEN} = 412, n_{DFP} = 375; Risk-increasing circumstances: n_{PEN} = 412, n_{DFP} = 375).

[a] PEN (Penitentiary, Freiburg), DFP (Department of Forensic Psychiatry, Munich)

*$p < .05$. ** $p < .001$.

3.4 Discussion

The methodological minimum requirements[4], published by an interdisciplinary working group and labeled as "recommendations" as early as 2006, were primarily intended for forensic experts, but also for judges, public prosecutors and defense attorneys. They were intended to facilitate criminal risk assessment and its evaluation of their validity. The recommendations do not constitute binding legal criteria, consequently not meeting them does not constitute a failure of law. It is explicitly pointed out that experts can deviate from the criteria if there are objective reasons for doing so. The recommendations are also to assist in delimiting which questions were to be the subject of the expert assessment and which questions would have to be answered exclusively by the court alone. The revised recommendations for risk assessment reports (Boetticher et al., 2019, Kröber et al. 2019) also list that there is no claim to binding force, but that the expert would have to explain if he or she deviated from the principles for professional reasons in individual cases. Even if the minimum standards or recommendations have in the meantime gained general recognition in the professional sciences, they are by no means considered by every forensic expert (Böhm, 2018; Wertz & Kury, 2017a; 2017b; Wertz et al., 2018). Insofar as the context of the present study, it cannot necessarily be said that there was a lack of quality if the orderer and expert did not follow the requirements. Nevertheless, it is already clear from the formulation of the methodological minimum requirements or recommendations that the processing of these questions should be regularly implemented, and thus no particularly high-threshold quality standards are described. The minimum level of differentiation in the judicial order is described, from which deviations should only be made in justified exceptional cases.

[4] In the recently published update of the minimum requirements for risk assessment reports, these are now also referred to as "recommendations" in the title of the paper (Boetticher et al., 2019; Kröber et al., 2019).

The results of the present study revealed significant group differences in judicial orders concerning the prognostic questions (1), (2) and (4) before and after the publication of the minimum requirements. This suggests a stronger orientation of the judicial orders towards the minimum requirements or recommendations across institutions, from which it can be concluded that the published quality standards influenced the judicial ordering practice in the direction of an increasing differentiation of the questions due to the published recommendations. However, exclusively considering the judicial questions posed to the experts after 2006, the differentiation of the formulated orders, even after the publication of the minimum requirements, is still heterogeneous. Overall, there was a stronger orientation of the judicial orders towards the questions specified in the minimum requirements, but the subsequent practice of judicial ordering continued to be heterogeneous and did not show any consistent prognostic questions. This was verified again when analyzing the wording of the judicial orders of the university department in Munich. Thus, hardly any consistent formulations of prognostic questions could be determined in the present sample.

The empirical examination of the experts' responses in assessment reports revealed significant group differences for the overall sample, at least regarding the responses to fundamental prognostic questions (1) and (2) before and after the publication of the methodological minimum requirements. However, the separate consideration of the institutional sub-samples showed that there was a statistically significant increase in frequency of responses to questions (1), (2), and (4) after 2006 in reports of the university department, while the external reports of the penitentiary in Freiburg did not show significantly more frequent or comprehensive responses over time with respect to any of the four fundamental prognostic questions. The results in favour of the university department can be explained in part by the institution's collaboration on the minimum requirements and the presumably more pronounced research proximity of the university department, so that university institutions in some respects represent a special status in expert report practice.

Considering only the sub-sample of the external reports of the penitentiary in Freiburg, even after the publication of the minimum requirements, answers to the fundamental questions were only provided in comparatively few reports. The current risk assessment practice by external experts at a penitentiary thus continued to be heterogeneous and did not provide consistent answers to the fundamental prognostic questions. The results indicate that the publication of the minimum requirements or the recommendations for risk assessment reports were more likely to be considered in the university institution than in general external expert practice. It can therefore be assumed for the future that research recommendations, contrary to the efforts of the initiators, will not necessarily lead to the desired effect across institutions and in a timely manner.

Regardless of the institutional background, significant correlations were found between the consideration of the prognostic questions in judicial orders and the answers given by the experts in assessment reports. The results of the present study show that the judicial order for a risk assessment report has an influence on the answering of the fundamental prognostic questions and can thus significantly increase the quality of risk assessments in the sense of a better usability in the proceedings. Thus, the judicial orders represent important quality assurance potentials. The published recommendations for risk assessments should therefore not only be implemented by experts but also considered by judicial orderers.

4. Quality of Criminal Responsibility Assessment in Clinical Practice – Minimum Requirements, Risk Assessment and Consideration in Judicial Verdicts[5]

Abstract

Assessments of the criminal responsibility of individuals convicted of sexual and/or violent offenses are of public interest, not least due to their influence on sentencing decisions regarding placement in a forensic hospital. The research literature points to a heterogeneous quality of these assessments in practice. In 2007, an interdisciplinary task force published minimum requirements for criminal responsibility reports. There is little empirical evidence on whether and how these requirements are put into practice. The current study examined the application of these minimum requirements, inspecting a sample of 230 expert reports provided by two departments of forensic psychiatry affiliated to the Psychiatric University Hospitals in Munich and Berlin. As about half of the analysed reports was delivered either before or after the publication of the requirements, we checked whether these revealed any impact on the reports' quality. In addition, we examined the court decisions regarding the consideration of the expert statements for a subsample of 130 cases providing this data. In summary, the implementation of the minimum requirements increased from period one to period two considered in this evaluation. However, risk assessments and the consideration of expert findings in court decisions continue to vary a lot in practice. On the one hand, the results indicate a (partial) positive effect of the introduction of minimum standards; on the other hand, more efforts are needed regarding quality assurance of criminal responsibility assessments.

Keywords: quality assurance, criminal responsibility assessment, minimum requirements, risk assessment, juridical verdicts

[5] Paper published as Wertz, M., Hausam, J., Konrad, N., Schiltz, K., Imhoff, R.& Rettenberger, M. (2021). Qualität von Schuldfähigkeitsgutachten – Mindestanforderungen, unterbringungsrelevante Gefährlichkeitsprognose und Berücksichtigung im richterlichen Urteil [Quality of criminal responsibility reports – Minimum requirements, risk assessment, and consideration in court decisions]. Recht & Psychiatrie, 39(4), 202 – 211.

4.1 Quality of criminal responsibility assessment reports

The criminal responsibility assessment is of particular importance due to the considerable consequences for the individuals assessed in the criminal proceedings. It has a decisive influence on sentencing decisions regarding placement in a forensic hospital and the duration of the deprivation of liberty (Müller & Nedopil, 2017; Prüter-Schwarte et al. 2019; Verrel, 1995). Criminal responsibility assessment reports frequently attract public interest and are regularly the subject of media discussions (Dahle et al., 2012; Kury & Obergfell-Fuchs 2012; Müller & Nedopil2017; Verrel, 2015). The scientific literature points to a heterogeneous quality of expert witness reports, which, in addition to formal deficiencies, also refer to aspects of content (Dahle et al. 2012; Fegert et al., 2003; Kunzl et al., 2009; Kunzl & Pfäfflin 2011; Schläfke et al., 2006; Schnoor, 2009). Regarding the legal side, it is also not uncommon for overturned sentences to be justified by errors in the criminal responsibility assessment or the assessment of the prerequisites of mandatory treatment in a forensic mental hospital (Mosbacher, 2020, p. 446). As a result of the ongoing discussion about the quality of expert witness reports and quality assurance, methodological minimum requirements for criminal responsibility assessments were published in 2007 (Boetticher et al., 2007). Empirical evidence on the extent to which these methodological minimum requirements are implemented in practice is still scarce. An empirical study conducted in Germany after the publication of the methodological minimum requirements still found urgent potential for improvement in almost 200 analysed criminal responsibility assessment reports (e.g., Prüter-Schwarte et al., 2019). While the formal methodological minimum requirements were largely implemented, "considerable weaknesses [...] were found in areas essential for the assessment of criminal responsibility in the reports examined, which question the results of the assessments" (Prüter-Schwarte et al., 2019, p. 207). Another study of 50 German criminal responsibility reports conducted in 2017 still showed formal inadequacies and deficiencies in content (Stübner et al., 2018).

4.1.1 Risk assessment in criminal responsibility assessment reports

In cases where ordering of mandatory treatment is being considered, experts are regularly asked for a criminal risk assessment in addition to the assessment of the prerequisites of the suspended or diminished criminal responsibility (Müller & Nedopil, 2017). A placement in a forensic psychiatric hospital may only be ordered if it is related to the significance of the offenses committed and those that can be further expected as well as the degree of dangerousness posed by the examinee (Kammeier, 2018). Criminal risk assessments are not only linked to the existence of a crime-relevant mental disorder (article 63 of the German penal code), a crime-relevant tendency towards intoxication (article 64 of the German penal code), or significant offenses (article 66 of the German penal code). Experts should also provide explanations on the probability, nature, and severity of future offenses as well as risk-altering measures and circumstances (Boetticher et al., 2019). In this respect, judicial questions about the prerequisites of a mandatory treatment require the expert to deal with the risk of recidivism in addition to criminal responsibility. So far, it hardly seems to be empirically recorded how the risk assessment – i.e., the judicial question regarding the risk of significant offenses expected in the future as a prerequisite for the ordering of a mandatory treatment – is implemented in criminal responsibility reports. While it has been shown that the published methodological minimum requirements have led to an improvement in the quality of criminal risk assessments (Wertz et al., 2018; 2020), it remains unclear whether a stronger orientation towards the published recommendations can also be found in the context of risk assessments in criminal responsibility assessment reports.

As shown in current studies, the overturning of sentences due to defective criminal responsibility assessment reports are oftentimes based on risk assessments in the context of the articles 63 and 64 of the German penal code. Since, according to case law, the decisive criterion is a high degree of probability for the commission of significantly unlawful acts, it is not sufficient from an expert point of view to assess whether such acts are "to be expected,"

"possible," "not improbable," or even "not excludable" (Mosbacher, 2020, p. 446ff). The prerequisites of mandatory treatment primarily deal with one's individual dangerousness due to mental disorders, which can only be assumed if there is a justified risk of significantly unlawful acts.

The increasing use of standardized risk assessment instruments has been noted in the expert context (Dahle & Lehmann, 2018; Leygraf, 2015; Müller & Nedopil, 2017; Rettenberger, 2018; Rettenberger & von Franqué, 2013), in the context of social therapeutic institutions (Etzler & Rettenberger, 2019), in the practice of outpatient aftercare for individuals convicted of sexual offenses (Gregório Hertz et al., 2019), as well as in the practices of granting privileges in forensic psychiatric hospitals (Sklenarova et al., 2020). The added value of standardized risk assessment instruments is also emphasized in the methodological minimum requirements or recommendations for risk assessment reports (Boetticher et al., 2019; Kroeber et al., 2019). However, yet no empirical data are available on the extent to which this development can also be found in criminal responsibility assessment reports in general practice.

4.1.2 Consideration of criminal responsibility assessment reports in judicial verdicts

Although the methodological minimum requirements for criminal responsibility assessment reports were primarily intended for forensic experts, they should also provide support for judges, public prosecutors and defense attorneys. They should help assess the validity of the reports and uncover dubious expertise and contradictions between experts (Boetticher et al., 2007). The minimum requirements are intended, among other things, to contribute to a dialogue capability between orderers and experts (Pfister, 2019). Therefore, as a first step, it seems relevant in terms of quality assurance to consider how transparently the prerequisites of the suspended or diminished criminal responsibility are discussed in assessment reports and, second, to examine the extent to which expert recommendations are adopted by the courts and reported in the verdict. Previous studies (Schläfke et al., 2006;

Schnoor, 2009) suggest that the courts largely adopt the results of the expert reports on criminal responsibility, and usually all recommendations are incorporated verbatim into the verdict without further discussion. A critical examination of the expert recommendations or a discussion of these was only rarely present (Schläfke et al., 2006; Schnoor, 2009; Verrel, 2015). The review of judicial adoption behaviour after the publication of the methodological minimum requirements showed a relation to the proportion of fully met minimum criteria at a similarly high adoption rate. At the same time, however, there is still a predominantly formulaic to absent judicial discussion of the expert reports (Schöttle et al., 2013).

4.1.3 Aim of the study

The aim of the present study was to examine whether the publication of methodological minimum requirements (Boetticher et al., 2007) has led to an increased implementation of those in criminal responsibility assessment reports in clinical practice. Furthermore, we analysed how transparently the prerequisites of the suspended or diminished criminal responsibility are discussed in assessment reports, how the criminal risk assessment required by the text of the law as a necessary prerequisite for mandatory treatment according to articles 63, 64, and 66 of the German penal code is implemented in criminal responsibility assessment reports, whether published recommendations for criminal risk assessment reports have led to a stronger orientation towards these in the risk assessment in criminal responsibility assessment reports, and the form in which the expert findings are considered in in judicial verdicts.

4.2 Method

4.2.1 Sample

The analysis units of the present study were n = 230 criminal responsibility assessment reports from two independent institutions: n = 100 assessment reports from the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of the Ludwig-Maximilian University in Munich and n = 130 assessment reports from the Institute

of Forensic Psychiatry of the Charité in Berlin. While in the university department in Munich a random selection of reports was gathered, in the Charité in Berlin a full survey of corresponding criminal responsibility reports from the time periods in question was conducted. Reports that were based solely on records were excluded. The assessment reports were divided into a pre- and post-group regarding the time of publication of the minimum requirements (Boetticher et al., 2007); 109 reports (47.4%) were prepared between 1990 and 2006 (pre) and 121 reports (52.6%) were prepared between 2007 and 2016 (post).
The majority of examinees were assessed by psychiatric experts (*n* = 225, 97.8%). In Berlin, five reports were conducted by psychologists. 65 (65.0%) of the one hundred reports from the university department in Munich, and 42 (32.3%) of the 130 reports from Charité in Berlin, were conducted interdisciplinary (by consulting an additional psychological test examination report).

On average, the examinees at the university department in Munich were just under 38 years old (*M* = 37.81, *SD* = 13.84) and those at the Charité in Berlin just under 36 years old (*M* = 35.77, *SD* = 14.06). Twelve percent of the sample from the university department in Munich (*n* = 12) and seven assessment reports from the Charité in Berlin (5.4%) referred to female individuals. Thirty-four subjects at the university department in Munich (43.0%) had at least one prior conviction at the time of the study according to the Federal Central Register (*M* = 5.29 prior convictions, *SD* = 9.32). At the Charité in Berlin, 66.2% of the examinees had a criminal record (*n* = 86; *M* = 4.42, *SD* = 5.55).

4.2.2 Data collection and material

The criminal responsibility assessment reports and the requested court procedural outcomes were systematically recorded by two student employees using a survey questionnaire. Both surveyors were trained in advance on the use of the survey questionnaire. To validate the quality of the ratings, interrater reliability with an experienced scientific expert was determined on a randomly selected sample of expert reports. The analyses for the

implementation of the methodological minimum requirements based on a sample of 13 assessment reports from the university department in Munich resulted in reliability coefficients that can be classified as very high according to the usual standard (Leonhart, 2004) for the total score for the implementation of the methodological minimum, *ICC* = .853, $p < .001$, 95% CI [.52, .96].

4.2.3 Implementation of methodological minimum requirements

Based on the methodological minimum requirements for criminal responsibility assessment reports, a survey instrument was developed comprising a total of 47 items. The general part is divided into formal (15 item) and content-related (10 items) minimum requirements, according analogously to the published minimum requirements. Sum scores were formed for the formal (value range of 0–23) and content-related (0–20 points) minimum requirements as well as a total score (0–43). A higher score in each case indicated a stronger implementation of the methodological minimum requirements.

In order to examine the "core content-related questions" of the criminal responsibility assessments (Prüter-Schwarte et al., 2019), the presence of a discussion regarding the prerequisites of the articles 20, 21 of the German penal code (classification to legally defined mental disorders, capacity to understand and control own behaviour) was gathered dichotomously. A mere mention of the expert's final assessment judgement was not sufficient to an affirmative answer.

4.2.4 Implementation of risk assessment in criminal responsibility assessment reports

The criminal risk assessment was examined with regard to the scope, the answering of the fundamental prognostic questions according to current recommendations for risk assessment reports (Boetticher et al., 2019), and the use of risk assessment instruments. The expert's response to the four fundamental prognostic questions was rated on the basis of a three-level scaling (see also Wertz et al., 2020). The use of risk assessment instruments was

recorded dichotomously ("yes"/"no") as well as in terms of methodological approach. A distinction was made between second- and third-generation actuarial and clinical-structured risk assessment instruments in terms of "structured professional judgment" (SPJ)[6] (Rettenberger, 2018).

4.2.5 Consideration in judicial verdicts

Information on the outcome of the respective proceedings was requested in writing from the ordering public prosecutors' offices and courts. This information was available for a subsample (n = 136). First, the extent to which the assessment reports were mentioned in the judicial verdicts and the exact wording of the judicial verdict were recorded. In addition, the final judicial decisions were gathered with regard to the classification of legally defined mental disorders as well as the assessment of the capability to understand and control behaviour and were compared with the expert findings from the written assessment reports.

4.3 Results

Table 4.1 presents the judicial order, the index offense, the psychiatric diagnoses according to ICD-10, and the expert's responses to criminal responsibility and the prerequisites of mandatory treatment of the total sample of included criminal responsibility assessment reports depending on institution and time period of contribution. In more than two-thirds of the total sample of reports, the judicial order explicitly included the assessment of the prerequisites of a mandatory treatment, while in just under one-third, only the criminal responsibility of the examinees was questioned (according to articles 20, 21 of the German penal code). Almost 64% of the reports were conducted on individuals convicted or charged of sexual and/or violent offenses, while the other offenses contained property and fraud offenses, offenses by narcotics law, and other offenses. No mental disorder was diagnosed in

[6] The Psychopathy Checklist - Revised (PCL-R; Hare, 2003), frequently used in expert reports, is a diagnostic assessment instrument for the clinical construct of *psychopathy*, a robust predictor of persistent delinquency. Although it is not a classic risk assessment instrument, it was considered as SPJ instrument because of its methodological approach.

approximately 30% of the sample. Personality and sexual preference disorders accounted for just under a quarter of the diagnoses made, followed by schizophrenia, schizotypal and delusional disorders (F20-29, 17.8%), and mental and behavioural disorders caused by psychotropic substances (F10-19, 15.7%). In just under half of the reports, the diagnoses could not be classified by the legally defined mental disorders according to article 20 of the German penal code. The most frequent classifications accordingto the legal definitions were severe mental diseases (36.1%) and other severe disturbances of the mind (13.0%). While more than 85% of the assessment reports indicated an undiminished capability to understand own behaviour, almost half of the reports indicated a suspended or diminished capability to control own behaviour. Overall, just over 50% of the examinees were classified as criminally responsible. In over 65% of the sample, the prerequisites of mandatory treatment in a forensic psychiatric hospital were not met. In total, there were no statistically significant differences between the groups (pre/post) in the report- and examinee-related characteristics. No differences were found between the institutions either (see Table 4.1), which is why the total sample of reports without differentiation by institution was used for the further analyses.

4.3.1 Implementation of methodological minimum requirements

Table 4.2 shows the implementation of the formal and content-related methodological minimum requirements for both groups (before and after the publication of minimum requirements). All items had higher mean scores in the post group, suggesting greater implementation of the methodological minimum requirements over time: the groups differed significantly, each with a large effect for the formal, $t(156.93) = -8.52$, $p < .001$, $d = 1.16$, 95% CI [-1.44, -0.88] and content-related methodological minimum requirements, $t(209.84) = -3.98$, $p < .001$, $d = 0.53$, 95% CI [-0.79, -0.27], as well as for the total score, $t(196.92) = -6.26$, $p < .001$, $d = 0.84$, 95% CI [-1.11, -0.57].

Table 4.1

Statistical Sample Description of Criminal Responsibility Reports depending on Institution and Time Period of Contribution (N = 230)

Variable	LMU, Munich (*n* = 100)				Charité, Berlin (*n* = 130)				Total sample (*N* = 230)			
	pre (*n* = 50)		post (*n* = 50)		pre (*n* = 59)		post (*n* =71)		pre (*n* =109)		post (*n* =121)	
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
Judicial Order												
Articles 20/21	13	26.0	13	26.0	30	50.8	10	14.1	43	39.4	23	19.0
Articles 20/21/63	9	18.0	6	12.0	5	8.5	14	19.7	14	12.8	20	16.5
Articles 20/21/64	2	4.0	6	12.0	3	5.1	4	5.6	5	4.7	10	8.3
Articles 20/21/63/64	26	52.0	25	50.0	14	23.7	32	45.1	40	36.7	57	47.1
Articles 20/21/66	0	0.0	0	0.0	1	1.7	2	2.8	1	0.9	2	1.7
Articles 20/21/63/64/66	0	0.0	0	0.0	6	10.2	9	12.7	6	5.5	9	7.4
Index Offense												
Violent Offense	27	54.0	19	38.0	29	49.2	32	45.1	56	51.4	51	42.1
Sexual Offense	3	6.0	5	10.0	6	10.2	13	18.3	9	8.3	18	14.9
Violent/Sexual Offense	4	8.0	2	4.0	4	6.7	3	4.2	8	7.3	5	4.1
Property/Fraud Offense	6	12.0	14	28.0	9	15.3	12	16.9	15	13.8	26	21.5
Offense by Narcotics Law	4	8.0	3	6.0	2	3.4	3	4.2	6	5.5	6	5.0
Other Offense	6	12.0	7	14.0	9	15.2	8	11.3	15	13.8	15	12.4
Psychiatric Diagnoses (ICD-10)												
No Diagnoses	18	36.0	10	20.0	23	39.0	19	26.8	41	37.6	29	24.0
F00-F09	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
F10-F19	10	20.0	10	20.0	7	11.9	9	12.7	17	15.6	19	15.7
F20-F29	10	20.0	10	20.0	11	18.6	10	14.1	21	19.3	20	16.4
F30-F39	0	0.0	5	10.0	2	3.4	1	1.4	2	1.8	6	5.0
F40-F49	0	0.0	0	0.0	2	3.4	2	2.8	2	1.8	2	1.7

Variable	LMU, Munich (n = 100)				Charité, Berlin (n = 130)				Total sample (N = 230)			
	pre (n = 50)		post (n = 50)		pre (n = 59)		post (n =71)		pre (n =109)		post (n =121)	
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
F50-F59	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
F60-F69	10	20.0	12	24.0	11	18.6	22	30.9	21	19.3	34	28.1
F70-F79	2	4.0	3	6.0	2	3.4	7	9.9	4	3.7	10	8.3
F80-F89	0	0.0	0	0.0	0	0.0	1	1.4	0	0.0	1	0.8
F90-F99	0	0.0	0	0.0	1	0.0	0	0.0	1	0.9	0	0.0
Legally Defined Mental Disorders												
No Classification	22	44.0	18	36.0	26	44.1	44	62.0	48	44.0	62	51.2
Severe Mental Disease	19	38.0	17	34.0	25	42.4	22	31.0	44	40.4	39	32.2
Severe Disturbances of Consciousness	1	2.0	1	2.0	0	0.0	0	0.0	1	0.9	1	0.8
Feeble Mindedness	1	2.0	0	0.0	2	3.3	2	2.8	3	2.8	2	1.7
Other Severe Disturbances of the Mind	7	14.0	14	28.0	6	10.2	3	4.2	13	11.9	17	14.1
Capability to understand behaviour												
capable	46	92.0	49	98.0	42	71.2	60	84.5	88	80.7	109	90.1
diminished capable	3	6.0	1	2.0	10	16.9	6	8.5	13	11.9	7	5.8
ambiguous[a]	1	2.0	0	0.0	7	11.9	5	7.0	8	7.4	5	4.1
Capability to control behaviour[b]												
capable	23	47.9	25	51.0	26	44.1	46	65.7	49	45.8	71	59.7
diminished capable	21	34.8	23	46.9	27	45.7	16	22.9	48	44.9	39	32.8
ambiguous[a]	4	8.3	1	2.1	6	10.2	8	11.4	10	9.3	9	7.5
Criminal Responsibility												
responsible	25	50.0	25	50.0	26	44.1	44	62.0	51	46.8	69	57.1
not responsible according to article 20	2	4.0	2	4.0	13	22.0	7	9.9	15	13.8	9	7.4
diminished responsible according to article 21	20	40.0	18	36.0	17	28.8	17	23.9	37	33.9	35	28.9
ambiguous[a]	3	6.0	5	10.0	3	5.1	3	4.2	6	5.5	8	6.6

Variable	LMU, Munich (n = 100)				Charité, Berlin (n = 130)				Total sample (N = 230)			
	pre (n = 50)		post (n = 50)		pre (n = 59)		post (n =71)		pre (n =109)		post (n =121)	
	n	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*	*n*	*%*
Prerequisites for Mandatory Treatment												
No Prerequisites met	28	56.0	36	72.0	45	76.2	41	57.7	73	66.9	77	63.6
Prerequisites according to article 63	10	20.0	7	14.0	7	11.9	11	15.5	17	15.7	18	14.9
Prerequisites according to article 64	9	18.0	4	8.0	1	1.7	10	14.1	10	9.2	14	11.6
Prerequisites according to article 66	0	0.0	0	0.0	2	3.4	7	9.9	2	1.8	7	5.8
ambiguous[a]	3	6.0	3	6.0	4	6.8	2	2.8	7	6.4	5	4.1

Note. pre/post (before and after the publication of minimum requirements for criminal responsibility assessment reports (Boetticher et al., 2007); LMU, Munich (Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of Ludwig-Maximilian University in Munich); Charité Berlin (Institute of Forensic Psychiatry of the Charité in Berlin)

[a] The assessment was not clearly possible from the expert(s') point of view (or not to be assumed, but also not to be excluded), which is why several alternative assessments were listed and left to the normative decision by the court. In most cases, reference was made to clarification in the main hearing.

[b] Since in some cases there was a diminished capability to understand behaviour, the examination of the capability to control behaviour was obsolete (the sample number thus varies slightly depending on the institution)

Table 4.2

Implementation of the Formal and Content-related Methodological Minimum Requirements for Criminal Responsibility Assessment Reports (N = 230)

Item	Designation	pre (*n* =109)		Post (*n* =121)	
		M	*SD*	*M*	*SD*
	Formal minimum requirements (0 - 23)	**19.28**	**2.28**	**21.35**	**1.17**
1	Orderer[a]	1.00	0.00	1.00	0.00
2	Judicial order	1.83	0.40	1.98	0.13
3	Location[a]	0.91	0.29	0.98	0.16
4	Date[a]	0.89	0.31	1.00	0.00
5	Time and scope of the assessment	1.06	0.49	1.20	0.42
6	Clarification[a]	0.84	0.36	0.97	0.18
7	Special assessment methods and documentation	1.91	0.32	1.97	0.22
8	Sources of knowledge	1.90	0.30	1.99	0.09
9	Separation of findings and interpretation	1.89	0.39	1.97	0.18
10	Separation of medical knowledge and subjective opinion	1.92	0.28	2.00	0.00
11	Ambiguities and difficulties	1.50	0.86	1.93	0.36
12	Involved experts and employees	1.98	0.19	2.00	0.00
13	Citation practice[a]	0.33	0.47	0.79	0.41
14	Structure[a]	0.92	0.28	0.99	0.09
15	Provisional nature[a]	0.39	0.49	0.59	0.49
	Conten-related minimum requirements (0 – 20)	**14.51**	**4.80**	**16.83**	**3.95**
16	Comprehensiveness of exploration	1.66	0.51	1.80	0.40
17	Assessment methods	1.88	0.35	1.93	0.26
18	Diagnostic system	1.23	0.82	1.47	0.71
19	Differential diagnostic considerations	1.29	0.79	1.68	0.57
20	General functional impairments	1.53	0.63	1.74	0.54
21	Functional impairments while offense	1.48	0.70	1.67	0.62
22	Classification to legally defined mental disorders	1.56	0.66	1.82	0.45

Item	Designation	pre (n =109)		Post (n =121)	
		M	*SD*	*M*	*SD*
23	Severity of the disorder	1.50	0.62	1.78	0.48
24	Offense-relevant impairment of the capability to understand and control behaviour	1.17	0.79	1.47	0.68
25	Alternative assessment	1.20	0.88	1.50	0.79
	Minimum requirements Total (0 - 43)	**33.79**	**6.00**	**38.18**	**4.42**

Note. pre/post = before/after publication of the minimum requirements (Boetticher et al. , 2007)

[a] dichotomously gathered items

4.3.2 Transparent discussion of the prerequisites of the articles 20, 21 of the German penal code

Subsequently, the discussion and substantiation of the prerequisites of the articles 20, 21 of the German penal code in the assessment reports were examined. For the total sample, there was a significant pre-post difference – starting from a high baseline level of over 90% – between the assessment reports before and after the publication of the methodological minimum requirements. Results showed an increasing discussion and substantiation of the classification to the legally defined mental disorders according to article 20 of the German penal code over time, with an effect size of *Cramér's V* = 0 .16 that can be classified as low. On the contrary, no more frequent (critical) discussion of the capability to understand and control own behaviour could be observed. Overall, in assessment reports before and after publication of the methodological minimum requirements, there was a predominant discussion of these capabilities in (just under) 75% or more of the total sample (cf. Table 4.3).

In addition, the examination of the verbatim formulations of the discussions in the assessment reports identified frequently used formulations. While the discussion of the classification of the expert findings to a legally defined mental disorder according to article 20 of the German penal code was strongly oriented toward differential diagnostic considerations

(with reference to underlying diagnostic systems) and the assessment of the severity of a disorder, the presentation of the functional impairments with differentiation between capability to understand and the capability to control own behaviour was mostly transparent and comprehensible but sometimes very different in detail. It ranged from one sentence to several pages in the written assessment report. The heterogeneity described above was also reflected in the level of detail in the written assessment reports; the total number of characters in the discussion sections varied considerably between the different written assessment reports.

Table 4.3

Absolute and Relative Distribution of the Existence of a Discussion of the Prerequisites of the Articles 20, 21 of the German Penal Code depending on the Time Period of Contribution (N = 230)

Discussion of the prerequisites of the articles 20, 21 of the German penal code	n_{pre} (%)	n_{post} (%)	$^2\chi(1)$	*p*	*V*
Classification to legally defined mental disorders	100 (91.7)	119 (98.3)	5.49	.019	0.16
Capability to understand behaviour	79 (72.5)	90 (74.4)	0.11	.744	0.02
Capability to control behaviour	92 (84.4)	101 (83.5)	0.04	.848	0.01

Note. pre (n = 109)/post (n = 121) (before and after publication of minimum requirements, Boetticher et al., 2007)

4.3.3 Implementation of risk assessment in criminal responsibility assessment reports

All reports in which the judicial orderer questioned a possible accommodation in a psychiatric hospital and in which, according to the expert, there was a limited capability to understand or control own behaviour, were analysed regarding the scope of the risk assessment, the answering of the fundamental prognostic questions and the use of risk

assessment instruments (n = 110)[7]. In the pre group, a criminal risk assessment was identified in 28 of 60 assessment reports (46.7%). In the post group, in 25 of 50 criminal responsibility assessment reports (50.0%) a criminal risk assessment was identified. The further analyses referred to this sub sample of 53 assessment reports. Regarding the scope of the risk assessments (in words), there was a higher number of words in the post group (M = 266.40 words, SD = 337.83) than in the pre group (M = 128.25 words, SD = 170.03), but this difference just missed statistical significance, $t(51) = -1.91$, $p = .062$, $d = 0.53$.

In these assessment reports, there was an overall lack of or incomplete response to the fundamental prognostic questions in well over half to just under three quarters of the assessment reports prior to the publication of the methodological minimum requirements. The extensive response to the question of risk-reducing measures was found in just over half of the assessment reports (see Table 4.4). Significant pre-post differences emerged in greater consideration of fundamental prognostic questions in assessment reports since 2007 regarding the frequency of mention of the (2) nature, frequency, and severity and (4) risk-increasing circumstances for the overall sample. Mention of the (1) probability and (3) risk-reducing measures did not differ over time. The effect sizes calculated for the strength of the difference varied from low values for (1) probability and for (3) risk-reducing measures to a medium difference value for (4) risk-increasing circumstances and for (2) nature, frequency, and severity (see Table 4.4).

Before 2007, risk assessment instruments were used in only one of 28 assessment reports (3.6%). After the publication of the minimum requirements, in nine of the 25 expert reports (36%) risk assessment instruments were used (cf. Table 4.5). The final risk communication was categorical in all expert reports with one exception, in which the risk was additionally communicated numerically.

[7] Assessment reports in which an accommodation to forensic psychiatric hospital according to article 64 of the German penal code was discussed irrespective of the capability to understand and control behaviour, were excluded.

Table 4.4

Answering to the Four Fundamental Prognostic Questions in Criminal Responsibility Assessment Reports including a Risk Assessment over Time (n = 53)

Answer	n_{pre}			n_{post}			$\chi^2(2)$	p	V
	Absent[a]	Incomplete[a]	Comprehensive[a]	Absent[a]	Incomplete[a]	Comprehensive[a]			
(1) Probability of re-offenses	3	16	9	3	10	12	1.65	.439	0.18
(2) Nature, frequency, severity	9	11	8	1	13	11	6.89	.032	0.36
(3) Risk-reducing measures	6	7	15	2	5	18	2.44	.295	0.22
(4) Risk-increasing circumstances	10	10	8	2	9	14	6.88	.032	0.36

Note. pre (n = 28)/ post (n = 25) (before and after publication of minimum requirements, cf. Boetticher et al., 2007)

[a] absent (no reference to the question), incomplete (only partial reference to the question), comprehensive (complete reference to the question)

Table 4.5

*Use of Risk Assessment Instruments in Criminal Responsibility Assessment Reports including a Risk Assessment over Time (*n *= 53)*

Use of risk assessment instruments	n_{pre}		n_{post}		$\chi^2(1)$	p	V
	yes	no	yes	No			
2nd generation	0	28	4	21	4.86	.028	0.30
3rd generation	0	28	3	22	3.56	.059	0.26
SPJ[a]	1	27	8	17	7.57	.006	0.38
Total	1	27	9	19	9.07	.003	0.41

Note. pre(n = 28)/ post (n = 25) (before and after publication of minimum requirements, cf. Boetticher et al., 2007)

[a] SPJ (Structured Professional Judgement)

4.3.4 Consideration in judicial verdicts

Table 4.6 shows a consideration of expert findings in judicial verdicts (or in the corresponding judicial documents available) in three quarters of the verdict documents obtained. The discussion of the prerequisites of the articles 20, 21 of the German penal code in the judicial verdicts showed a documentation of the classification to the legally defined mental disorders as well as the assessment of the capability to understand and control behaviour in about half of the examined verdicts or judicial decisions. Overall, there were no significant differences in the pre-post comparison between judicial verdicts before and after the publication of the methodological minimum requirements. Discrepancies between expert recommendations of the written assessment reports and the final judicial verdict decisions were found in ten of the 124 (8.1%) verdict documents examined. The discrepancies included assessments of the prerequisites of the articles 20, 21 of the German penal code.

In addition, numerous different formulations could be identified in the judicial verdicts, which – irrespective of the necessary orientation to different legal contexts and specifications – illustrated the heterogeneity of the judicial discussion of the prerequisites of the articles 20, 21 of the German penal code in judicial practice. The total number of characters in the discussion parts of the verdicts varied significantly between the different courts.

Table 4.6

Absolute and Relative Distribution of the Consideration of Expert Findings in Judicial Verdicts, the Presence of a Discussion of the Prerequisites of the Articles 20, 21 of the German Penal Code in the Judicial Verdicts, and the Discrepancies between Expert Recommendations and Final Verdict Decisions depending on the Time of Contribution (N = 136)

Variables	Pre				Post				Total sample				$\chi^2(1)$	*p*	*V*
	yes		No		yes		no		yes		no				
	n	%	n	%	n	%	n	%	n	%	n	%			
Consideration of expert findings in judicial verdicts	67	82.7	14	17.3	39	70.9	16	29.1	106	77.9	30	22.1	2.66	.103	0.14
Discussion of the prerequisites of articles 20,21, German penal code[a]															
Classification to legal mental disorders	39	50.6	38	49.4	19	38.8	30	61.2	58	46.0	68	54	1.7	.192	0.12
Capability to understand behaviour	40	51.9	37	48.1	17	34.7	32	65.3	57	45.2	69	54.8	3.6	.058	0.17
Capability to control behaviour	47	61.0	30	39.0	23	46.9	26	53.2	70	55.6	56	44.4	2.41	.12	0.14
Discrepancies between judicial verdicts and expert recommendations[b]	5	6.7	70	93.3	5	10.2	44	89.8	10	8.1	114	91.9	0.5	.479	0.06

Note. Pre (n = 81) /post (n = 55) (before and after publication of minimum requirements (Boetticher et al., 2007); $n_{\text{Classification to legal mental disorders}} = 126$; $n_{\text{Capability to understand behaviour}} = 126$; $n_{\text{Capability to control behaviour}} = 126$; $n_{\textit{Discrepancies between judicial verdicts and expert recommendation s}} = 126$; $N = 136$

[a] Ten court decisions could not be included in the analyses because they did not contain any information regarding the relevant question.

[b] Twelve court decisions could not be included in the analyses because they did not contain any information regarding the relevant question.

4.4 Discussion

The results regarding an implementation of the methodological minimum requirements showed significant differences over time. This indicates an improvement in the quality of expert assessment practice as a result of the publication of the methodological minimum requirements and is in line with comparable studies regarding criminal risk assessment (Wertz et al., 2018). In addition, the separate analysis of the implementation of the "core content-related criteria" (Prüter-Schwarte et al., 2019, p. 128) showed that the transparent and comprehensible discussion and substantiation of the classification of the legally defined mental disorders as well as the detailed discussion of the capability to understand and control behaviour – as required by the minimum requirements – were implemented in about 75% or more of the assessment reports. Significant differences between the assessment reports before and after the publication of the methodological minimum requirements were only evident in a more frequent discussion of the classification to the legally defined mental disorders according to article 20 of the German penal code, considering a correspondingly high initial level. Additionally, results regarding an examination of the verbatim formulations of the discussions illustrate the heterogeneity of the expert discussion concerning the prerequisites of the articles 20, 21 of the German penal code from a qualitative and quantitative perspective. Very different emphases could be identified regarding contextual substantiations and transparent presentations. The heterogeneity over time, despite the (at least partially) significantly stronger expert consideration of the prerequisites, was also reflected in the level of detail in the assessment reports.

The results thus provide indications of a desirable implementation of the methodological minimum requirements into practice and contrast with earlier studies in which implementation was assessed more critically (e.g., Prüter-Schwarte et al., 2019). Nonetheless, it is important to note, in a limiting manner, the special status of university institutions, whose heads were both members of the working group that developed the criteria catalogue (cf.

Boetticher et al., 2007). In university institutions, it can also be assumed that there is an increased institutional quality assurance (Wertz et al., 2018). Considering a limited comparability with older studies due to different methodological approaches (e.g., different operationalisation of the methodological minimum requirements), this could be an explanation for the more positive results within the present study.

Even if the methodological minimum requirements have gained general appreciation in research and practice in the meantime, they are by no means taken into account by every forensic expert (Dahle et al., 2012; Kunzl et al., 2009; Kunzl & Pfäfflin, 2011; Schnoor, 2009; Verell, 2015). It should be noted that the methodological minimum requirements do not represent binding legal criteria, the non-implementation of which would constitute a legal defect in any case. In this respect, it is not necessarily a quality defect in the context of the present study if experts do not follow the criteria. Nevertheless, formulation of the minimum requirements for example already illustrates that a discussion of the prerequisites of the criminal responsibility assessment as well as a correct answering of the questions should be conducted regularly. Thus, no particular high-quality standards are described. The criteria reflect the minimum level of differentiation in the assessment reports, from which discrepancies should only be made in factually substantiated exceptional cases. Conversely, compliance with the methodological minimum requirements does not guarantee a “correct” assessment and expert (Konrad, 2010), especially since the minimum requirements are also criticized alongside concerns with operationalization (Eisenberg, 2005) and clinical and practical applicability for assessing the severity of diverse mental disorder groups (Konrad, Huchzermeier & Rasch, 2019; Dobbrunz & Briken, 2020; Fuß et al., 2020).

Regarding this critical discussion, a further examination of the methodological minimum requirements seems useful. In addition to new empirical findings, changes in the legal framework could also be considered and mapped. Similarly, the recommendations for

risk assessment reports have recently been updated (Boetticher et al., 2019; Kroeber et al., 2019).

The criminal risk assessment in criminal responsibility assessment reports in which the orderer questioned the prerequisites of a mandatory treatment and in which the expert stated a limited capability to understand or control behaviour showed a heterogeneous implementation from a quantitative and qualitative point of view. Only half of the reports in which a risk assessment would have been indicated contained a risk assessment. The scope of the assessment in the written report varied greatly. There was also incomplete answering to the fundamental prognostic questions and sporadic use of standardized risk assessment instruments. There were no significant changes over time. The study results indicate a need for action for quality assurance in this area, which has already been emphatically criticized in the past (Schmitt & Rettenberger, 2015). In this respect, it seems obvious that research findings from this area should also be accounted for to a greater extent in the risk assessment for possible placement to mandatory treatment in criminal responsibility assessment reports. Conclusions are also supported by the recently updated recommendations for risk assessment reports (Boetticher et al., 2019; Kroeber et al., 2019).

Finally, the examination of the information on procedural outcomes revealed the heterogeneity of judicial consideration of expert findings. The limits of expert witness report evaluations in judicial verdicts were in line with a previous study in which the scope of the entire responsibility verdict was comparatively small, averaging less than two pages and accounting for well under one percent of the total verdict length (Verrel, 1995). Equal to the findings of the current study, Verrel (1995) showed that no discussion of the expert findings was found in verdicts in almost 20% of all cases with expert participation. In the majority of cases, the expert's findings were merely reproduced and evaluated in only ten percent of the cases. In the current study, there were no significant differences over time, which seems to

indicate that the methodological minimum requirements do not have a decisive influence on judicial practice.

The results show potential for quality assurance not only for the experts but also for the orderers of criminal responsibility assessment reports. The methodological minimum requirements for criminal responsibility assessment reports should not only be implemented by experts but should be considered by orderers – especially in the context of a transparent presentation of the prerequisites for mandatory treatment in verdicts. In terms of general considerations of quality assurance in the field of criminal responsibility assessment, the provision of transcripts of judicial verdicts according to article 475 (4) of the German penal code could be helpful for experts to improve their ability to communicate psycho-scientific findings to the court regarding knowledge of the consideration of the assessment report in verdicts (Pfister, 2019).

Nevertheless, the present study shows methodological limitations. Above all, some criteria of the methodological minimum requirements were difficult to assess retrospectively based on the written assessment reports. Furthermore, when evaluating the outcome of the proceedings, it must also be taken into account that the complete judicial verdict was not always available; in some cases only abridged versions of the verdict, decisions, or copies of the verdict could be obtained. In addition, it was usually not possible to determine from the verdicts if the expert deviated from the preliminary written assessment report during oral testimony at the main hearing. The present study examined the minimum requirements and thus only one of many possible measures for quality assurance of assessment reports. A final assessment of quality was not the aim of this study and would exceed the limits of an operationalized set of criteria for methodological minimum requirements. For these purposes, other measures of quality assurance must be pursued. Only recently, for example, the peer review process known in science was applied to expert witness reports in family law

(Kannegießer, 2018). A similar quality-assurance procedure would also be conceivable and desirable for criminal responsibility reports (Banse, 2017).

Irrespective of these limitations, the present study identified potential for improvement for both experts and orderers in terms of standardized practice in the assessment of criminal responsibility. Noticing an increased compliance with formal and content-related criteria in assessment reports due to the publication of the methodological minimum requirements, can be encouraging regarding an even more efficient implementation of minimum requirements. Although the heterogeneity of methodological implementation and the level of detail in the assessment reports were more apparent, also improvements in the implementation of risk assessments were identified over time.

5. The Use and Reporting Practice of Psychological Tests in German Risk and Criminal Responsibility Expert Reports[8]

Abstract

Assessment reports about individuals charged/convicted of offenses have an influence on significant personal consequences for examinees by sentencing decisions regarding placement in a forensic hospital or prison. As there is evidence that unstructured clinical judgments have limited accuracy, research-based practice recommendations call for the use of standardized measures and for experts to base their assessments on empirically supported psychological tests. Previous findings on the actual use of psychometric tests indicate an increasing but still heterogeneous use of psychological tests, highlighting the continued relevance of a professional debate on best diagnostic practice. A potential shortcoming, however, is that these studies almost exclusively relied on clinicians' self-reports. The present paper presents an analysis based on the actual (retrospectively assessed) usage in risk (n = 489) and criminal responsibility assessment reports (n = 272) as well as corresponding psychological test reports (n = 313) for examinees in Germany between 1990 and 2016. In accordance with previous survey data, results showed a frequent usage of a diverse range of psychological tests. Contrarily, performance-based personality tests, typically subsumed under so-called (semi-) projective personality tests, are still implemented regularly in forensic-clinical practice, although number and frequency seem to be decreasing. Taken together, the findings offer an important assessment of forensic psychological testing in diagnostic practice.

Keywords: psychological test, forensic assessment, risk assessment, criminal responsibility, expert reports

[8] Paper published as Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R. & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. Psychology, Crime & Law. Advance online publication. https://doi.org/10.1080/1068316X.2022.2063286

5.1 Introduction

5.1.1 Psychological Testing in Forensic Settings

A plethora of research has repeatedly established the superiority of statistical, actuarial, or mechanical prediction methods and the limited accuracy of unstructured clinical judgments not only for criminal risk assessments but for a diverse range of aspects of human behavior (Bengtson & Långström, 2007; Grove & Meehl, 1996; McLaughlin & Kan, 2014; Mehl, 2013; Rettenberger & Eher, 2016). Because a structured clinical approach is also indispensable in forensic assessments, clinical conclusions about relevant aspects of personality, cognitive functions, intelligence, or mental disorders of examinees in assessment reports should follow strict scientific standards, also including the use of standardized psychological measures based on scientific theory and technique. Due to psychometric characteristics of psychological tests, relevant aspects can be measured more objectively, reliably, and validly by implementing a distinct degree of standardisation and formalisation, especially in terms of standard values and norm data. Besides clinical interviews, risk assessment instruments and file information, experts therefore consequently base their assessments on empirically supported psychological tests to gain psychometrically sound foundation of reports (American Psychological Association, 2013; McLaughlin & Kan, 2014; Miller, 2013). Hence, psychological assessment is a defining area of practice, training, and research for professional psychologists, a large majority believe psychological assessment is a valuable aid in assisting diagnostic and treatment decisions or in screening for cognitive or neuropsychological deficits (Wright et al., 2017).

A range of studies has reported that psychological testing is applied in a variety of different types of forensic evaluations, including risk and criminal responsibility assessments (Archer et al., 2006; Borum & Grisso, 1995; Lally, 2003; McLaughlin & Kan, 2014; Neal et al., 2019; Varela & Conroy, 2012; Viljoen et al., 2010; Watkins et al., 1995; Wright et al., 2017; Zapf et al., 2004). Different studies indicated that within forensic psychology and

psychiatry, psychological testing seems to be a commonly used source of information in expert reports (e.g., Cutler & Kovera, 2011; McLaughlin & Kan, 2014; Neal & Grisso, 2014; Serafim et al., 2015; Wright et al., 2017). Previously published surveys reported that among forensic psychologists, approximately 30% of the working time in forensic-clinical practice is spent in psychological testing (Archer et al., 2006). A more recent survey reported an average of 24% of direct practice time conducting any psychological assessment, while participants working in a forensic setting spent a significantly greater percentage of their time with psychological testing than those in non-forensic institutions (Wright et al., 2017). Although surveys consistently identified psychological testing as an important component of forensic evaluations, frequency, nature, and concrete application practice of tests and therefore degree of formalisation vary significantly by setting, legal question, and evaluation focus (Archer et al., 2006; Fuger et al., 2014; Lally, 2003; McLaughlin & Kan, 2014; Neal, 2018).

Due to survey data studying the frequency and acceptability of psychological tests in different areas of forensic psychology (Archer et al., 2006; Boccaccini & Brodsky, 1999; De Clerq & Vander Laenen, 2019; Lally, 2003; McLaughlin & Kan, 2014; Neal, 2018; Neal et al., 2019; Neal & Grisso, 2014; Viljoen et al., 2010; Wright et al., 2017), multiscale personality tests (like the *Minnesota Multiphase Personality Inventory-2* [MMPI-2; Engel et al., 2000], the *Personality Assessment Inventory* [PAI; Morey, 1991] or the *Revised NEO Personality Inventory* [NEO-PI-R; Costa & McCrae, 2008]) were generally mentioned most frequently. Consistent with the results of previous surveys, findings underline the continuing popularity of traditional clinical assessment instruments in forensic psychology, such as the MMPI-2 (Archer et al., 2006; Lally, 2003; Mclaughlin & Kan, 2014; Neal et al., 2019; Wright et al., 2017). Rated as suitable to examine the mental state at the offence in terms of cognitive and intellectual functions, various forms of the *Wechsler Adult Intelligence Scale-III* (WAIS-III; Wechsler, 1997) were recommended consistently by survey respondents, followed by other performance tests (Archer et al., 2006; Boccaccini & Brodsky, 1999; Lally, 2003; Neal

& Grisso, 2014; Viljoen et al., 2010; Wright et al., 2017). Another of the most commonly used group of assessment instruments were the symptom specific measures or inventories like the *Beck Depression Inventory* (BDI; Beck et al., 1979), as reported by a recent survey (Wright et al., 2017).

Contrary to the findings of Lally (2003) and McLaughlin and Kan (2014) - indicating that performance-based personality tests, typically classified as *projective* measurements (Meyer & Kurtz, 2006; Viglione & Rivera, 2003), were neither accepted nor were the least often used in practice - and despite science-based criteria for the admissibility, the predominant number of surveys and examinations of court-ordered expert reports suggest that multiple performance-based personality tests (like the *Rorschach test* [Rorschach, 1927] or the *Thematic Apperception Test* [Morgan & Murray, 1935]) continue to be used by a substantial number of psychologists (Archer et al., 2006; De Clerq & Vander Laenen, 2019; Neal et al., 2019; Viljoen et al., 2010). Taken together, international surveys reported that most forensic evaluations are using multiple psychological tests, noting the extreme variety regarding the number of tools used (De Clerq & Vander Laenen, 2019; Fuger et al., 2014; Lees-Haley et al., 1996; McLaughlin & Kan, 2014; Neal & Grisso, 2014; Warren et al., 2006; Wright et al., 2017).

However, despite these studies, other authors have criticized the fact that the knowledge of forensic psychological testing in practice is still lacking compared with general clinical practice (Lally, 2003; Neal et al., 2019). This means that, on the one hand, there are some comparatively frequently used and relatively popular measures like the MMPI-2 or the Wechsler Intelligence scales but, on the other hand, there is no consensus about psychological test usage and the degree of formalisation in forensic evaluations (Archer et al., 2006; Golden & Lashley, 2014; Gowensmith & McCallum, 2019; Lally, 2003; Richards et al., 2015). Even if studies exist that survey forensic psychologists about the tests they are using regularly, the knowledge about forensic psychological testing in forensic-clinical practice is still sparse

compared to other (i.e., non-forensic) clinical settings (Archer et al., 2006; Lally, 2003). For example, there is still little empirical data about reporting and consideration of psychological test results in forensic court reports.

5.1.2 The Legal Background of Expert Witnesses in the German Penal Law

In Germany, the majority of criminal assessment reports focus on the culpability and recidivism risk of individuals charged or convicted of sexual and/or violent offenses (Dunn et al., 2014). Reports are required to assess whether there is a clinical diagnosis which abolishes or diminishes the insight of the alleged person regarding the wrongfulness of the criminal act or the capability to act through a lack of insight, and if the offender poses a risk of future offences (due to his or her assumed mental disorder). The German legal system has three categories of culpability in mentally disordered offenders: culpable, diminished culpable, and not culpable (articles 20 and 21 of the German Penal Code). The forensic assessment requires a two-step evaluation: first, the alleged person must have a legally defined disorder at the time of the offense. These respective legal terms are: Severe Mental Disease, Severe Disturbances of Consciousness, Mental Retardation, and Other Severe Disturbances of the Mind (translations as close as possible to the legal code terminology; for an empirical evaluation of the stigmatizing nature of these terms see Rösch et al., 2021). Mental Diseases in the sense of the law are, for example, organic or genetic disorders, diseases of the brain, severe acute intoxications, and major mental disorders. The term Severe Disturbances of Consciousness means short but exceptionally severe mental disturbances in individuals, who are otherwise mentally sane (i.e., severely affective exceptional situations). The term refers to the evaluation of criminal responsibility regarding so-called affect crimes due to consciousness disturbances. Mental retardation is understood as a severe deficit of intelligence, which restricts substantially the ability to understand and act due to legal standards. All other mental disorders, like personality disorders, sexual deviations, and other chronic or longer lasting reactive disorders are classified with the term Other Severe Disturbances of the Mind.

If the defendant suffered from one of these disorders, the expert witness has to assess the functional consequences of the disorder at the moment of the criminal act. The alleged person, who was unable to understand the unlawfulness of the criminal act or lacked the ability to control his actions because of one of the four disorders mentioned in article 20, is usually considered not responsible. When his or her ability to control his or her actions was severely diminished, the responsibility is regarded as diminished (article 21 of the penal code). Finally, psychologists and psychiatrists will also be asked about the dangerousness of a defendant. Defendants who are not culpable and who are at the same time considered to be dangerous are sentenced to mandatory treatment in a forensic mental hospital. Those who are diminished responsible are sentenced to both hospital treatment and prison, and the time spent in prison will be deducted from the period of the prison sentence. Individuals suffering from addiction disorders are sentenced by article 64 of the German penal code and individuals suffering from any other diagnoses by article 63 of the German penal code. Furthermore, Germany also has provisions for preventive detention of persons regarded as dangerous serial offenders (article 66 of the German penal code). Risk assessments from mental health professionals will be required not only at the time of sentencing but several times thereafter, because detaining and releasing mentally ill or chronically dangerous offenders from hospitals or prisons depend strongly on the results of their risk assessment reports. The mental health professionals do not make decisions about detaining or releasing a person but give their advice to the courts, which then have to base their decisions upon their legal understanding and the empirical evidence of the individual case (Dunn et al., 2014).

Psychological test expertise is either part of a forensic expert report or forensic psychologists are asked to conduct an additional specific psychological test report for other mental health professionals like, for example, forensic psychiatrists (Rieger & Stadtland, 2013; Schneider et al., 2015; Wertz & Tippelt, 2019). Generally, the frequency of the usage of psychological testing within court statements is increasing (Habermeyer, 2008). Although

psychological testing seems to have a general acceptance within the forensic field in Germany (Habermeyer, 2008; Schneider et al., 2015), empirical findings about the usage and application of tests and especially about the consideration of forensic psychological testing in the final judgments of risk or criminal responsibility of individuals charged or convicted of offenses is still sparse. A German study found an increased usage of test instruments of 55% in 227 forensic evaluations regarding the preventive detention in comparison to reported 20% in a period of ten years (Habermeyer, 2008; Kinzig, 1997a, 1997b). A survey of the diagnostic practice in social therapy units in the German prison system (Etzler & Rettenberger, 2019) showed that psychological testing plays a key role in planning and monitoring treatment but at the same time indicates a relatively heterogenous diagnostic practice in these institutions. Another survey described the outpatient aftercare of individuals convicted of sexual offenses after release from prison in Germany, indicating a regular use of standardized diagnostic tests, mostly multidimensional personality instruments (Gregório Hertz et al., 2019).

As “Specialty Guidelines for Forensic Psychology” (American Psychological Association, 2013), a “Practice Guideline for the Forensic Assessment” by the American Academy of Psychiatry and the Law (Glancy et al., 2015), or “Best Practices in Forensic Mental Health Assessment” (Heilbrun et al., 2009) were developed and published internationally, a German assessment system of psychological tests provides additional evaluation criteria for usage of psychological testing, also including description of type, content and diagnostical purpose as well as norm sample, reliability, and validity of tests (Diagnostik- und Testkuratorium, 2018). In Germany, methodological minimum requirements were published for criminal responsibility (Boetticher et al., 2007) and risk assessment reports (Boetticher et al., 2006, 2019; Kröber et al., 2019) by an interdisciplinary working group, including recommendations for usage of psychological testing. According to these requirements, practical use and selection of psychological tests are based on psychometric

quality criteria validated by empirical studies. Tests also have to meet requirements according to adequate norm samples of the German-speaking area.

5.1.3 Study Objectives

It is thus still largely unknown whether the increase in self-reported usage of psychological tests is mirrored in a corresponding increase of the actual usage in forensic practice (as shown for the usage of risk assessment instruments in German risk assessment reports; Wertz et al., 2018). Furthermore, data about reporting practice and influence of psychological testing on corresponding assessments is lacking. The present study will remedy this research desideratum by providing empirical data about the use of psychological testing in criminal responsibility and risk assessment reports in Germany. More precisely, study objectives were to examine

- the number, frequency, content, and type of tests used in assessment reports in real practice,
- if frequency of test usage is increasing and dependent on type of forensic assessment (risk assessment vs. criminal responsibility), and
- the reporting practice and influence of test results on assessment reports (especially with regard to the final judgments).

While in previous studies the survey data were usually based on the self-reports of clinicians (e.g., Archer et al., 2006; Lally, 2003; Wright et al., 2017), in the present investigation a comprehensive amount of court reports about risk assessment and criminal responsibility cases were empirically analysed. Thereby, the analysis units were not self-reports of expert witnesses but assessment reports.

5.2 Methods

To examine the clinical relevance of forensic psychological testing in criminal risk and responsibility assessment, a retrospective analysis of risk and criminal responsibility assessment reports (N = 761) was conducted. The forensic evaluations underwent an

empirical validation regarding the concrete usage of psychological tests depending on the type of forensic assessment and the year of contribution. In assessment reports including psychological testing, content subgroup and type of test instruments were examined, and all identified used tests were listed. Data was systematically gathered; only data concerning accommodation (in forensic hospital) or imprisonment (in penitentiary) in criminal responsibility assessments was not available at the time of data collection because at the moment when the expert report was written, it was unclear whether the charged individual will serve a prison sentence (in cases where criminal responsibility was confirmed) or be transferred to a forensic psychiatric institution (in cases where criminal responsibility was denied).

5.2.1 Sample Description and Descriptive Data

The sample consisted of risk (n = 489) and criminal responsibility assessment reports (n = 272) for individuals charged or convicted of offenses which have been gathered from different German institutions representing common general forensic practice in Germany: the penitentiary in Freiburg (n = 150), the Institute of Forensic Psychiatry at the Charité, Berlin (n = 128), and the Department of Forensic Psychiatry of the University Hospital Munich (n = 483)[9]. Assessment reports conducted between 1990 and 2016 were ordered by diverse judicial parties to the proceedings like local or district courts, courts for the execution of prison, higher regional courts, and public prosecutors. All available risk assessment reports which

[9] The university hospital institutions in Berlin and Munich provide external forensic assessment reports for a diverse number of different courts or public prosecutors as well as education and training of forensic students, psychologists, psychiatrists, and other clinicians working in the forensic field. The penitentiary in Freiburg is run by the federal state Baden-Wuerttemberg in the southwest of Germany and has the primary goals of imprisonment and rehabilitation of prisoners. At a penitentiary, assessment reports are not done by the institution itself, but are mostly obtained by external, residential psychiatric or psychological experts, who are in most cases not affiliated to a specific academic or scientific institution. Overall, the institutions represent common forensic-clinical practice, as psychiatric hospitals and penitentiaries are not just asked regularly but in the majority of cases to conduct assessment reports in Germany.

were accessibly archived in institutions were included in the total sample. Given that we intended to gain data of adult male offenders, we excluded female and underage offenders. Reports that were only based on records without personal examination of individuals, as well as incompletely archived reports, were also excluded.

Overall, 110 different experts wrote the risk assessment reports. Nearly all risk reports were contributed by psychiatrists (92.2 %, *n* = 451), while 7.8 % (*n* = 38) were reported by psychologists. Three hundred seventy-six (76.9 %) criminal risk reports were reported by interdisciplinary teams (e.g., by including an additional psychological test report or an additional interview about the social rehabilitation process by a qualified sociologist). The descriptive statistics (see Table 5.1) present the number of included assessment reports depending on the type of forensic assessment, judicial order, accommodation (in forensic hospital) or imprisonment (in penitentiary), index offense, and psychiatric diagnoses of individuals charged or convicted of offenses. As at the University Hospital Munich psychological testing in forensic evaluations is reported via additional psychological test reports by default, a subsample of these additional psychological test reports (*n* = 313) was identified.

Table 5.1

Descriptive Statistics of all included Risk and Criminal Responsibility Assessment Reports

Variable	Criminal risk assessment reports		Criminal responsibility assessment reports	
	n	%	*n*	%
Accommodation/Imprisonment				
Penitentiary	159	32.5	[b]No data available	
Preventive Detention (§66[a])	106	21.7		
Forensic Psychiatry (§63[a])	214	43.8		
Forensic Psychiatry (§64[a])	10	2.0		
Judicial Order				
Granting Privileges	26	5.3	--	--
Release	235	48.1	--	--
Privileges/Release	110	22.5	--	--
Preconditions for §66[a]	26	5.3	--	--
Privileges of §66[a]	20	4.1	--	--
Release of §66[a]	56	11.5	--	--
Privileges/Release of §66[a]	16	3.3	--	--
§ 20/21[a]	--	--	70	25.7
§ 20/21/63[a]	--	--	43	15.8
§ 20/21/64[a]	--	--	18	6.6
§ 20/21/63/64[a]	--	--	116	42.6
§ 20/21/63/64/66[a]	--	--	25	9.2
Index Offence				
Violent Offence	212	43.4	129	47.4
Sexual Offence	75	15.3	35	12.9
Violent-/Sexual Offence	149	30.5	16	5.9
Property-/Fraud Offence	24	4.9	44	16.2
Offence by narcotics law	17	3.5	18	6.6
Arson	12	2.5	0	0
Other Offences	0	0.0	30	11
Psychiatric Diagnoses (ICD-10)[c]				
No Diagnoses	130	26.6	89	32.7
F00-F09	24	4.9	0	0.0
F10-F19	44	9	43	15.8
F20-F29	62	12.7	41	15.1
F30-F39	1	0.2	10	3.7
F40-F49	0	0.0	5	1.8
F50-F59	0	0.0	0	0.0
F60-F69	211	43.1	68	25
F70-F79	8	1.6	13	4.8
F80-F89	1	0.2	1	0.4
F90-F99	8	1.6	2	0.7

Notes. *N* = 761 (*n* = 489 risk assessment and *n* = 272 criminal responsibility reports*)*

[a] articles 20 (not culpable), 21 (diminished culpable), 63 (forensic psychiatry, individuals suffering from

any other diagnosis except addiction disorders), 64 (forensic psychiatry, individuals suffering from addiction disorders) or 66 (provisions for preventive detention of persons regarded as dangerous serial offenders) of the German penal code; see further explanations in detail in the section about the legal background of expert witnesses in the German penal law

[b] No data about accommodation/imprisonment (e.g., pretrial detention or as per criminal procedure code 126a) of individuals charged or convicted of offenses at time of criminal responsibility assessment was available by retrospective analysis of reports.

[c] F00-F09 = Organic, including symptomatic, mental disorders; F10-F19 = Mental and behavioural disorders due to psychoactive substance use; F20-F29 = Schizophrenia, schizotypal and delusional disorders; F30-F39 = Mood [affective] disorders; F40-F49 = Neurotic, stress-related and somatoform disorders; F50-F059 = Behavioural syndromes associated with physiological disturbances and physical factors; F60-F69 = Disorders of adult personality and behaviour; F70-F79 = Mental retardation; F80-F89 = Disorders of psychological development; F90-F99 = Behavioural and emotional disorders with onset usually occurring in childhood and adolescence

5.2.2 Empirical Data Collection and Procedure

Reports that were only based on records without personal examination of individuals, as well as incompletely archived reports, were excluded (n = 11). The number and frequency of used instruments was analysed, and afterwards the used tests were allocated to subcategories regarding their content or main measurement purpose (intelligence, specific cognitive functions, personality [multidimensional, specific, and performance-based[10]], assessment of mental disorders, psychosexual constructs, and symptom / performance validity). Additionally, the type of tests was categorized as performance-based personality measures, performance tests, questionnaires, and semi-structured interviews. Finally, the number and frequency of each used test according to legal question (risk assessment vs. criminal responsibility) was examined.

[10] Typically, and historically classified as (semi-)projective measurements (Meyer & Kurtz, 2006; Viglione & Rivera, 2003), in this study referred to as performance-based personality measures.

To examine the reporting practice of psychological testing in assessment reports, a sample of additional psychological test reports of the University Hospital Munich was identified (n = 313) and included in the analysis. Corresponding assessment reports underwent an empirical validation with respect to reporting of tests and test results of additional psychological test reports. Characteristics of reporting were rated on a three-point Likert-scale from “not at all“ (no reporting), “incomplete“ (reporting, but incomplete for all used test instruments) to “complete” reporting (reporting of all used instruments). To determine the interrater reliability of this rating, a random sample of 31 psychological test reports (10.1%) was co-rated by a trained second rater, which showed high reliability coefficients for the three rating categories of reporting characteristics with ICCs from .896 to .943, $p < .001$, 95% CIs [.801-.979].

5.3 Results

5.3.1 Use of Forensic Psychological Testing in Real Practice

Table 5.2 provides information on the use of test instruments depending on the type of forensic assessment (risk assessment vs. criminal responsibility). Findings show a significant difference in test usage in accordance to the type of forensic assessment (*Cramérs V* = .29): Three hundred sixty-five (74.6%) risk assessment reports included psychological tests, and in 123 (45.2%) criminal responsibility assessment reports, tests were used. Also, risk assessment reports on average included significantly more tests than criminal responsibility reports, $t(486) = 3.811$, $p < .001$, $d = .40$. Significant mean differences of test usage regarding the year of contribution could only be found in risk assessment reports. Results show more frequent use of psychological testing over time, while in criminal responsibility reports no significant differences could be identified (see Table 2).

Regardless of the type of assessment, personality tests were applied most, especially performance-based and multidimensional personality tests, whereas the most used types of tests were questionnaires. In criminal responsibility assessment reports, a higher percentage of

tests regarding cognitive functions and especially intelligence as well as more performance tests were used than in risk assessment reports. A separate analysis showed that the use and number of performance-based personality measures decreased significantly over time, while other types of test instruments showed a significant increase by year of contribution (see Table 5.2).

Table 5.2

Descriptive percentage Distribution of Test Usage in the Total Sample and (Number of) Used Test Instruments depending on Year of Contribution, Content Subgroup and Type of Test Instrument in Assessment Reports

Variable	Criminal risk assessment reports		Criminal responsibility assessment reports	
	N	*%*	*n*	*%*
Usage of psychological Tests[a]	365	74.6	123	45.1
Number of tests (mean, SD)[b]	7.2 (*SD*=3.1)		5.9 (*SD*=2.7)	
Year of Contribution[c]				
1990 - 1999	0[e]	0[e]	17[e]	53.1[e]
2000 - 2005	101	68.7	50	46.7
2006 - 2010	160	74.4	19	42.2
2011 - 2016	104	81.9	37	41.6
Content Subgroup				
Personality (total)	354	96.9	102	82.9
multidimensional	315	86.3	97	78.9
specific Traits	152	41.6	35	28.5
performance-based[d]	294	80.5	76	61.8
Cognitive Functions	159	43.6	65	52.8
Intelligence	153	41.9	99	80.5
Mental Disorder	96	26.3	17	13.8
Psychosexual Constructs	79	21.6	9	7.3
Symptom / Performance Validity	91	24.9	15	12.2
Type of Test				
Questionnaire	347	95.1	101	82.1
Performance-Based Personality Measures	325	89.0	84	68.3
Performance Test	265	72.6	105	85.4
Structured Interview	14	3.8	5	4.1

Notes. N = 761 (n = 489 risk assessment and n = 272 criminal responsibility reports)

[a] Significant difference regarding type of forensic assessment (criminal risk (M = 7.2; SD = 3.1) vs. criminal responsibility (M = 5.9; SD = 2.7) reports; $\chi^2(1) = 65.767$, $p < .001$, $V = .294$)

[b] Significant difference regarding type of forensic assessment (criminal risk vs. criminal responsibility

reports; $t(486) = 3.811$, $p < .001$, $d = .397$)

[c] Differences regarding year of contribution in criminal risk ($n = 365$), $\chi^2(2) = 6.266$, $p = .044$, $\varphi = .113$) and criminal responsibility reports ($n = 123$), $\chi^2(3) = 1.545$, $p = .672$, $\varphi = .075$. Additionally, use and number of projective measures decreased significantly over time ($\chi^2(2) = 6.843$, $p = .039$, $\varphi = .131$), while other types of test instruments showed a significant increase by year of contribution.

[d] Typically and historically classified as (semi-)projective measurements (Meyer & Kurtz, 2006), in this study referred to as performance-based personality measures

[e] Percentages in parentheses refer to relative number of reports with psychological testing in all conducted assessment reports in that time period

As reported in Table 5.3, generally a diverse range of tests was applied. In 365 risk assessment reports, 135 different tests were used. The three most frequently used psychological tests were performance-based personality measures (*Photo-Hand-Test*[11], *Rorschach-Test*, and the *Rosenzweig Picture-Frustration-Test*[12]), followed by multidimensional personality tests (MMPI-2, Personality Style and Disorder Inventory) and an intelligence test (*Wechsler Adult Intelligence Scale*). In the 123 criminal responsibility reports, 101 different tests were used: The *Wechsler Adult Intelligence Scale* was used most, followed by performance-based measures (*Photo-Hand-Test*, *Rosenzweig Picture-Frustration-Test*), and a multidimensional personality test (MMPI-2). In general, in the subsample of risk assessment reports, there was a focus on describing personality traits, while in criminal responsibility assessment reports, measuring intelligence was most frequent.

[11] The *Photo Hand Test* (FHT) is a diagnostic Measure aimed to measure aggressive dispositions. It uses photographs of hand gestures which test persons are required to interpret (Belschner et al., 1971).

[12] The *Rosenzweig Picture-Frustration Test* (PFT) is a traditionally so called semi-projective method used to measure frustration tolerance. It consists of 24 drawings which present a frustrating situation which test persons are required to interpret by identifying with and answering as the frustrated person in a dialogue (Hörmann & Moog, 1957).

Table 5.3

Percentage Distribution of Tests Used in Risk and in Criminal Responsibility Assessment Reports

Usage in percentage	Used test instruments in assessment reports including psychological testing	
	Criminal risk assessment reports[a]	Criminal responsibility assessment reports[b]
> 60%	--	--
50-60%	Photo-Hand-Test (58.9%) Rosenzweig Picture Frustration Test (52.1%)	Wechsler Adult Intelligence Scale (52.0%)
40-50%	Rorschach-Test (41.4%)	--
30-40%	Minnesota Multiphasic Personality Inventory®-2 (35.3%) Wechsler Adult Intelligence Scale (34.3%) Personality Style and Disorder Inventory (33.2%)	Rosenzweig-Picture-Frustration Test (38.1%) Minnesota Multiphasic Personality Inventory®-2 (35.6%) Photo-Hand-Test (34.0%)
20-30%	Drawing Completion Test (24.9%) Supernormality Scale (21.4%) The Thematic Apperception Test (20.6%)	Rorschach-Test (29.2%) Standard Progressive Matrices (24.3%) Personality Style and Disorder Inventory (23.5%) Drawing Completion Test (22.7%) Freiburg Personality Inventory-R (21.9%) The Thematic Apperception Test (20.3%)
10-20%	Standard Progressive Matrices (16.4%) Questionnaire for Identifying Factors of Aggression (16.4%) Basic system for measuring dementia (15.1%) Eysenck Personality Inventory (15.1%) Structured Inventory of Malingered Symptomatology (15.1%) Freiburg Personality Inventory-R (13.4%) Narcissism -Inventory (13.4%) Multiphasic Sex Inventory (12.6%) Short Cognitive Performance Test (11.8%) Inventory of clinical personality accentuation Supplementary Catalogue (10.1%)	d2-attention endurance test (17.0%) Benton-Test (16.2%) Narcissism -Inventory (12.2%) Giessen Test (11.3%) Basic System for Measuring Dementia (10.5%) Cerebral Insufficiency-Test (10.5%) Questionnaire for Identifying Factors of Aggression (10.5%)

Notes. N = 488 (n = 365 risk assessment and n =123 criminal responsibility reports

[a] 5-10% in risk assessment reports: Questionnaire on Beliefs in Competency and Control (9.9%), 16-Personality Factor Test - R (9.9%), Psychopathic Personality Inventory-Revised (9.3%), Cerebral Insufficiency-Test (9.3%), Implicit association test for sex offenders (8.2%), Giessen Test Self-Rating (8.0%), Child Identification Scale - Revised (7.4%), Vocabulary Test (7.4%), Inventory of clinical personality accentuations Basic inventory (7.1%), Wechsler Adult Intelligence Scale (7.1%), Scales of rape myth acceptance (6.9%),

Coping with stress Questionnaire (6.3%), Trail Making Test (6.3%), Benton-Test (5.8%), Bumby Child Molest Scale (5.8%), Eppendorfer Schizophrenia Inventory (5.5%), Spontaneous self-description (5.5%), Supplementary list of questions (5.2%); <5%: *n* = 97 other test instruments

[b] 5-10% in criminal responsibility reports: Supernormality Scale (8.9%), Eysenck Personality Inventory (8.1%), Multiple Choice Vocabulary Intelligence Test (7.3%), Freiburg Personality Inventory-R (6.5%) Psychopathic Personality Inventory-Revised (6.5%), Short Cognitive Performance Test (6.5%), Structured Inventory of Malingered Symptomatology (6.5%), <5%: *n* = 77 other test instruments

5.3.2 Reporting Practice of Forensic Psychological Testing

Results presented above underline the acceptance of forensic psychological testing in forensic-clinical practice, but they do not offer information on the relevance of this data source for final judgments. Therefore, reporting practice and the consideration of used tests and test results in the court reports were examined in the next step. A subsample of additional psychological test reports of the University Hospital Munich was identified (n = 313) and included in the analysis.

Information about the reporting practice of psychological testing (conducted in additional psychological test reports at the University Hospital Munich) is presented separately for risk assessment (*n* = 248) and criminal responsibility reports (*n* = 65). Table 5.4 provides information about the reporting of test instruments and results of psychological testing in assessment reports. While name, type of used test instruments, and concluding results of psychological testing were mostly reported in three-quarters of analysed reports, behavioural observation, (sub)scales, standard values, and total scale values were reported completely only in less than 10% of assessments. No significant differences between reporting of test instruments and results in criminal risk and responsibility assessment reports could be found.

Table 5.4

Percentage Distribution of Reporting of Psychological Test Results in Assessment Reports

Characteristics of Reporting	"not at all"		"incomplete"		"complete"	
	N	*%*	*n*	*%*	*n*	*%*
Risk Assessments						
Name/Type of Test Instruments	21	8.5	29	11.7	198	79.8
Behavioural Observation	126	50.8	101	40.7	21	8.5
(Sub)Scale/Standard Values	207	83.5	37	14.9	4	1.6
Total Scale Values	103	41.5	127	51.2	18	7.3
Concluding results of psychological testing	37	14.9	100	40.3	111	44.8
Criminal responsibility Assessments						
Name/Type of Test Instruments	2	3.1	16	24.6	47	72.3
Behavioural Observation	34	52.3	20	30.8	11	16.9
(Sub)Scale/Standard Values	44	67.7	17	26.2	4	6.2
Total Scale Values	13	20.0	34	52.3	18	27.7
Concluding results of psychological testing	2	3.1	30	46.2	33	50.7

Notes. N = 313 (n = 248 risk assessment and n = 65 criminal responsibility reports)

To examine if results of psychological testing are not just reported generally in risk assessment reports but are considered as relevant to answer the guiding forensic question, final judgments regarding risk of individuals charged or convicted of violent and sexual offenses in assessment reports referring to the corresponding, identified subsample of additional psychological test reports at the University Hospital (n = 248) were analysed. In 131 (52.8 %) final judgments of risk assessment reports, psychological test results were considered and discussed in terms of risk and protective factors regarding criminal risk of individuals charged or convicted of offenses. Particularly, test results about personality and cognitive functions of the examinee were taken into consideration.

5.4 Discussion

5.4.1 Usage of Forensic Psychological Testing

According to survey data based on clinicians' self-reports (e.g., Archer et al., 2006; Lally, 2003; Wright et al., 2017), it could be derived from the results of the present study that

clinical and forensic assessment instruments play a crucial role in forensic evaluations (Archer et al., 2006; Wright et al., 2017). A majority of professional psychologists consider psychological assessment as a valuable source of information which assists diagnostic or treatment decisions and the examination of cognitive and neuropsychological deficits. Interestingly, participants working in a forensic setting spent a significantly greater percentage of their time with psychological testing than those of non-forensic institutions (Wright et al., 2017). As with previous studies, the survey data was usually based on the self-reports of clinicians (e.g., Archer et al., 2006; Lally, 2003; Wright et al., 2017); the present investigation provides empirical data about the real-world use of psychological testing in criminal responsibility and risk assessment reports. The study results show a frequent usage of a diverse range of psychological tests in forensic-clinical practice, confirming the general acceptance within the forensic field. In risk assessment reports, psychological tests (mostly addressing personality traits) were used significantly more often (and in a significantly higher number) than in criminal responsibility assessment reports (primarily using intelligence measures). While the usage of psychological testing increased significantly in risk assessment reports over time consistent with previous studies (Neal & Grisso, 2014; Serafim et al., 2015; Wright et al., 2017), no significant difference could be confirmed regarding criminal responsibility reports. This is all the more noteworthy, as criminal responsibility assessments already had a lower level of test usage to begin with and it is legally mandatory to diagnose a psychological condition in a first step. However, in this German sample, psychological testing seemed to be more common and furthermore increasing in risk assessment reports than in criminal responsibility reports in recent years, documenting a growing relevance of interdisciplinarity in risk assessment in general.

In criminal responsibility assessment reports, more tests regarding cognitive functions and intelligence and more performance tests were used than in risk reports. In accordance with survey studies (Archer et al., 2006; Lally, 2003; Wright et al., 2017), the most often used

instruments were intelligence tests (especially *Wechsler Adult Intelligence Scale*), and personality tests (*Freiburg Personality Inventory-R,* MMPI-2). Previous survey studies had suggested that historically so-called *projective* measurements, in this study referred to as performance-based personality measures, were not accepted for usage in practice, which was in line with a vigorous debate concerning their acceptability in forensic evaluations (Lally, 2003; McLaughlin & Kan, 2014; Areh et al., 2021). A recent review concludes that the Rorschach test - as an example of projective or performance-based personality tests - does not meet proposed standards in terms of standardization and it should not be considered by psychologists in legal proceedings (Areh et al., 2021). Despite this criticism, surveys of professional test usage have documented that nevertheless, the Rorschach is one of the most frequently used clinical assessment instruments in forensic assessment reports (Archer et al., 2006; Borum & Grisso, 1995). Some authors (e.g. Meyer & Kurtz, 2006; Viglione & Rivera, 2003) argue that these types of performance-based personality tests like the Rorschach or Thematic Apperception Test are incorrectly equated as “unstructured” or “unstandardized”. Thus, these tests can also be used in a nomothetic (using normative data for interpretation) rather than just in an ideographic (merely based on clinical judgment) manner. For example, the recently developed *Rorschach Performance Assessment System* (R-PAS; Meyer et al., 2011) is considered to be extremely structured in terms of administration rules, coding instructions and interpretation based on normative data. In the present study, performance-based personality tests - predominantly used without normative data - were still implemented regularly, although number and frequency seemed to be decreasing over time.

Researchers, clinicians, and parties to proceeding still controversially question the admissibility of forensic psychological testing, not only regarding performance-based personality tests (Golden & Lashley, 2014; Gowensmith & McCallum, 2019; Lally, 2003; Schneider et al., 2015). Especially, the use of personality questionnaires is questioned in terms of positive response distortion, socially desirable responses, and inadequate norm samples for

forensic populations (Golden & Lashley, 2014; Heilbrun et al., 2007; Neal et al., 2019; Schneider et al., 2015). Therefore, German methodological minimum requirements were published for criminal responsibility (Boetticher et al., 2007) and risk assessment reports (Boetticher et al., 2006, 2019; Kröber et al., 2019), including recommending practical use of psychological tests based on psychometric quality criteria validated by empirical studies and adequate norm samples of German-speaking areas. Other authors (Heilbrun et al., 2009; McLaughlin & Kan, 2014) argue that even when using a specific instrument is inappropriate due to, for example, a missing norm sample, the use of a structured approach is still important in that evaluators should "structure" their assessments through standardized psychological testing. However, present results of forensic psychological testing in forensic-clinical practice show regardless of type of assessment, that personality tests were applied most, especially performance-based, and multidimensional ones. In terms of response distortion, symptom validity tests and performance validity tests are relevant to forensic evaluations. Assessment reports should include an assessment of response style, malingering and performance validity (McLaughlin & Kan, 2014; Heilbrun et al., 2009). As symptom validity tests were regularly considered in the sample, performance validity tests were not used more often than in 5% of assessment reports. Given the high prevalence of the use of performance tests, particularly in criminal responsibility reports, it can be seen as a noteworthy finding in German practice of forensic psychological testing.

All in all, the present results provide empirical data for the wide acceptance as well as variety and heterogeneity of test usage in forensic evaluations, as there seems to be no consensus about psychological test usage and the degree of formalisation in forensic evaluations (as reported in Archer et al., 2006; Golden & Lashley, 2014; Gowensmith & McCallum, 2019; Lally, 2003; Richards et al., 2015). Furthermore, the knowledge about forensic psychological testing in forensic-clinical practice is still sparse compared to other (i.e., nonforensic) clinical settings (Archer et al., 2006; Lally, 2003).

5.4.2 Reporting practice of forensic psychological testing

Apart from selection and application of test instruments in psychological test reports, a transparent and verifiable reporting of test instruments and results in corresponding assessment reports is seen as an essential requirement for forensic evaluations (Boetticher et al., 2006, 2007, 2019; Diagnostik- und Testkuratorium, 2018; Kröber et al., 2019). Regarding the ongoing debate about the importance of this data source for the final judgments, information about consideration of psychological testing and results especially in (psychiatric or psychological) final judgments of risk of individuals is lacking. Results show that name, type of used test instruments, and concluding results of psychological testing were mostly reported in three quarters of analysed reports, underlining integration and importance of psychological testing in risk and criminal responsibility assessments in real practice. In more than a half of final conclusions of risk assessment reports, psychological test results were not just reported generally but even considered and discussed in terms of risk and protective factors regarding criminal risk of individuals charged or convicted of offenses. Especially test results about personality and cognitive functions of the persons charged or convicted of offenses were taken into consideration. Supporting previous survey data, results underline the importance of psychological testing in forensic evaluations in real practice, especially for risk assessment.

5.4.3 Limitations

In interpreting these findings, it is important to note that most assessment reports were conducted at a University Hospital. Therefore, the standards and capacities of inclusion of psychological testing described may be different to those of clinicians who are not members of scientific forensic organizations like university hospitals or other research institutions. However, the integrated institutions represent common general forensic-clinical practice in Germany. Also, criminal responsibility reports were almost exclusively contributed by psychiatrists, which could have caused a lower frequency of usage of test instruments due to

profession and knowledge about psychological testing. In line with Warren et al. (2006) but contrary to Fuger et al., (2014), external witness experts significantly integrated or ordered psychological testing more often than psychiatrists. Furthermore, additional psychological test reports were contributed by a small number of experts, so individual preferences and opinions of experts e.g., on validity of projective measures had an influence on present results. Taken together, the results indicate the need for further discussion about inclusion and heterogeneity of psychological testing in the field of forensic psychiatry and psychology, including more research on non-scientific institutions. Furthermore, another research desideratum is the examination of the quality of assessment reports related to psychological test findings. As the frequency of psychological test use is only one part of the information necessary for understanding recent practice, future research should also aim to explore rationales for use of tests to form their decisions.

5.4.4 Conclusion

All in all, psychological testing can serve as a source of information to confirm hypotheses about psychological constructs relevant to the legal question. In accordance with previous survey data, results showed a generally frequent usage of a diverse range of psychological tests. While in previous studies the survey data was usually based on the self-reports of clinicians (e.g., Archer et al., 2006; Lally, 2003; Wright et al., 2017), the present study provides empirical data about the use of psychological testing in criminal responsibility and risk assessment reports in Germany. Taken together, the findings offer an important assessment of forensic psychological testing in diagnostic practice and underline the variety and heterogeneity of test usage and wide acceptance of the use of psychological tests in forensic evaluations. Results indicate not only the importance of forensic psychological testing in risk and criminal responsibility reports but also underline the need for further research especially on the application and reporting of forensic psychological test usage in real practice in contrast to survey data.

6. The Use of Standardized Risk Assessment Instruments in Germany: Empirical Results about Frequency and Risk Communication Depending on Report- and Offender-Related Variables[13]

Abstract

In German criminal law, risk assessment represents an essential component in penitentiary and forensic psychiatric hospital settings. The scientific literature discusses different methodological approaches for criminal risk assessment, which may be divided into clinical-intuitive, statistical-actuarial, structured-professional, and clinical-idiographic predictions. Previous study results illustrated the advantages of standardized prediction methods compared to intuitive and unstructured judgments and pointed to the significantly higher predictive validity of standardized risk assessment methods. In the present study, the use of actuarial and structured-professional risk assessment methods and instruments of N = 605 expert witness reports written between 1999 and 2016 was analyzed regarding different report- (time, institutional context, expert profession) and offender-related characteristics (index delinquency, diagnosis, incarceration or placement). Despite the increasing use of actuarial and structured-professional risk assessment instruments over time, a relatively heterogeneous application practice of risk assessessment methods was found. Furthermore, differences were identified regarding both report- and offender-related characteristics. While the results indicate an increasing standardization of risk assessment in pratice, there still seems to be a need for further quality assurance efforts.

Keywords: risk assessment, actuarial prediction, statistical prediction, structured professional judgment (SPJ), clinical prediction.

[13] Paper published as Wertz, M., & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale [The use of standardized risk assessment instruments in the practice of risk assessment: Empirical findings on frequency and risk communication as a function of assessment- and subject-related characteristics.]. Forensische Psychiatrie und Psychotherapie, 28(3), 241-261.

6.1 Introduction

6.1.1 Relevance of Criminal Risk Assessments in German Criminal Law

Criminal risk assessments of individuals already convicted or charged of offenses are one of the most important tasks of legal decision-makers in German criminal law. Prognostic decisions entail far-reaching freedom-related consequences for individuals (Rettenberger, 2018; 2019), as they significantly influence the (duration of) imprisonment or accommodation in a psychiatric forensic hospital (Müller & Nedopil, 2017; Prüter-Schwarte et al., 2019; Verrel 1995; 2015).

In the case of legal decisions of considerable consequence, experts from psychoscientific professions are regularly consulted (Boetticher et al., 2019) in order to ensure a scientifically sound basis for the legal decisions (Dahle, 2010). Risk assessments, in addition to assisting with potential granting of privileges or release measures for individuals convicted of offenses, also serve as a basis for assessing the prerequisites of a custodial measure in the context of criminal responsibility assessments. Criminal risk assessments thus represent an essential component of the areas of responsibility of persons working in penitentiaries or forensic psychiatry's or of external expert witnesses (Basdekis-Jozsa, et al., 2013; Dahle, 2005; Gretenkord, 2013). Prognostic decisions are also an important component of planning and implementing the most effective risk management measures possible (Rettenberger, 2019), the naming and discussion of which are explicitly called for in the current recommendations for risk assessment reports (Boetticher et al., 2019). In addition to general remarks on risk assessment and possible treatment options, the formulated requirements also address whether and how a persisting criminal risk can be controlled or reduced by appropriate risk management measures.

6.1.2 Methodological Approaches to Criminal Risk Assessment

Criminal risk assessment reports are a professional, methodological, and clinical challenge for experts (Rettenberger & Eher, 2016). They show an increasing complexity of

differentiated questions, which require the analysis of different data and diagnostic findings and their methodologically sound integration based on valid assessment models (Dahle & Lehmann, 2018). The scientific literature discusses different approaches for criminal risk assessment, which may be divided into clinical-intuitive, statistical-actuarial, structured-professional, and clinical-idiographic predictions, each with advantages and limitations (Rettenberger, 2018). While the so-called first generation of intuitive prediction, based only on the intuition and experience of the expert, does not represent scientifically sound expertise, statistical-actuarial instruments represent a standardized, empirically based assessment of static (2nd generation) and dynamic (3rd generation) risk factors. Here, risk variables correlating with recidivism are combined into point values in order to infer empirically determined probability values from these (Dahle, 2005; Hanson, 2009).

Checklists or criteria catalogues that serve to clinically structure risk and protective factors can be summarized under the methodological approach of structured clinical assessment instruments (so-called *Structured Professional Judgement* [SPJ]; von Franqué, 2013). In contrast to actuarial instruments, the item values are not added up to an overall value but are used in the sense of a clinical explanatory model, on which basis the individual risk of recidivism is to be structurally assessed. The combination of actuarial and SPJ instruments with a clinical individual case-based assessment method is often referred to as a - comparatively complex – clinical-idiographic approach (for a detailed discussion, cf. Rettenberger, 2019). While the 2nd and 3rd generation actuarial approaches imply a standardized and more empirically orientated approach based on group statistically derived findings, the degree of individual case-based orientation required by legislation (Boetticher et al., 2019; Kröber et al., 2019) increases with the use of SPJ instruments as well as clinical-idiographic approaches.

However, standardized risk assessment instruments always provide empirically based probability statements which are an essential aid to decision-making in the required individual

case-based assessment. A thorough and systematic risk assessment by statistical-actuarial approaches, with reference to group-statistical average correlations, forms a sound basis on which individual case-related hypotheses regarding possible risk areas can be derived, taking into account the broadest possible empirical knowledge (Dahle, 2010). The use of appropriate instruments serves to reliably determine baseline risk, identify care and treatment goals, and transparently present the course of treatment. Taken together, they can be considered *state of the art in* current risk assessment practice (Dahle, 2010; Rettenberger, 2019). The standardized instruments allow an assignment of relevant subgroups of offenders, on the basis of which valid statements about expected recidivism rates in terms of relative and absolute risk can be derived (Eher et al., 2019).

6.1.3 Superiority of Standardized Criminal Risk Assessment Approaches

Statistically based and standardized risk assessment instruments show a significantly higher predictive validity compared to judgments based on clinical intuition or experience (e.g., Rettenberger, 2018; 2019; Viljoen et al., 2021). The findings have been repeatedly and unambiguously confirmed, not just in general prognostic research but specifically in criminal risk assessment. Intuitive and experience-based risk assessment reports often do not show a predictive validity above chance, independent of data quality and professional experience of experts (Grove & Meehl, 1996; Grove et al., 2000; Hanson & Morton-Bourgon, 2009, Quinsey et al., 2006), and can thus be regarded as partly responsible for the heterogeneous quality in risk assessment reports (Rettenberger & Eher, 2016; Wertz et al., 2018). Despite a variety of scientifically based methods of criminal risk assessment (Rettenberger & von Franque, 2013), the intuitive and experience-based approach seems to remain prevalent in clinical practice (Haubner-Mclean et al., 2014).

According to these findings, numerous actuarial and clinical-structured risk assessment instruments for individuals already charged or convicted of offenses have been developed and examined for their psychometric criteria (Rettenberger, 2019). This has

contributed decisively to an increasing improvement in the predictive accuracy of risk assessments (Gretenkord, 2013; Rettenberger & Eher, 2012). The use of scientifically sound risk assessment instruments can therefore be considered an integral part of criminal risk assessments as well as for the planning and implementation of the most effective risk management measures possible. However, it does not seem possible to formulate generally applicable guidelines or recommendations for the use of specific risk assessment instruments, as this must always be oriented to the context and the specific question posed (Kröber et al., 2019; Rettenberger, 2019).

6.1.4 The Use of Statistical-Actuarial and Structured-Professional Risk Assessment Instruments in Practice

Despite numerous research papers on the reliability and (predictive) validity of standardized criminal risk assessments, comparatively few studies can be found that dealt with the use of actuarial and clinical-structured risk assessment instruments in practice. While at least some studies are available for the international, primarily Anglo-American area (e.g., Archer et al., 2006; Singh et al., 2014; Viljoen et al., 2010), only a few empirical studies have so far been conducted for the German-speaking area on the form in which criminal risk assessments are made in professional practice. There is particularly little evidence about the risk assessment practice, especially of external experts.

A survey study on criminal risk assessment practice in Germany (Rettenberger at al., 2017) showed that, according to their own statements, psychological and psychiatric experts use standardized risk assessment instruments in the majority of cases. Standardized risk assessment instruments were used in over 50% of all assessments, and in the period of the twelve months prior to the survey, practitioners reported using a risk assessment instrument in 65% of cases. The group of psychological experts used risk assessment instruments significantly more often than their psychiatric colleagues. The most frequently cited instruments were the *Psychopathy Checklist-Revised* (PCL-R; Hare, 2003), the *Historical*

Clinical Risk Management-20 (HCR-20; Müller-Isberner, et al., 1998), the *Forensic Operationalized Therapy Risk Evaluation System* (FOTRES; Rossegger et al., 2011), and the *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006). According to the survey, the use of actuarial and clinical-structured instruments in this context not only served to assess future anticipated offending but was also used for treatment planning and to review the course of treatment.

A survey of department heads and specialized services on intramural psychodiagnostic practice in social therapy facilities of penitentiaries (Etzler & Rettenberger, 2019) revealed that initial diagnostics were more standardized than ongoing and final diagnostic examinations. A total of 23 different risk assessment instruments were mentioned, in descending order of frequency of use: the HCR-20, the PCL-R, the *Level of Service Inventory-Revised* (LSI-R; Dahle et al., 2012), the *Sexual Violence Risk-20* (SVR-20; Müller-Ibserner et al., 2000), the VRAG, the *Static-99* (Harris et al., 2003), the *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006), the so-called *Dittmann List* (Dittmann, 2000), and the *Stable-2007* (Matthes & Rettenberger, 2008). These survey results also suggested that criminal risk assessment practice in social therapy settings is relatively heterogeneous.

In another survey study on criminal risk assessment practice in outpatient aftercare for individuals convicted or charged of sexual offenses in Germany (Gregório Hertz et al., 2019), more than three-quarters of participating institutions reported using criminal risk assessment instruments. The most frequently cited instruments were the Static-99, the Stable-2007, the PCL-R, the LSI-R, the HCR-20, the VRAG, the Dittmann List, and the *Structured Assessment of Protective Factors for Violence Risk* (SAPROF; de Vogel et al., 2011; Yoon et al., 2013), again showing the heterogeneity of the use of standardized risk assessment instruments.

6.1.5 Communication of Criminal Risk

There is relatively little evidence yet about how criminal risk is finally communicated in risk assessment reports to decision makers (de Vogel et al., 2020). However, the presentation and communication of risk assessment is essential for effective risk management and should ideally be transparent, comprehensible and unambiguous (Boetticher et al., 2019). Basically, a distinction can be made between a nominal and quantitative risk communication. Nominal risk communication involves a dichotomous or categorical weighting of risk, for example in the form of "low," "moderate," or "high" risk, which, however, may not be subject to any clear assignment to recidivism probabilities and thus to a significant extent to intuitive subjective evaluation criteria.

In numerical communication, probability or frequency information is expressed in the form of category-specific absolute recidivism rates for defined periods of time, usually requiring a reference value (e.g., in the form of recidivism base rates) for a content classification of the quantitative information (de Vogel et al., 2020; Eher et al., 2019; Gretenkord, 2013; Nedopil, 2005). The "5-Category Model for Sexual Offenders" provides a merging of relative and absolute risk based on transparent, comprehensible, combined nominal and quantitative risk communication orientated to the recidivism baseline rate (Eher et al., 2019), which was initially developed for the Static-99, but in the future will pe provided for further instruments. Other authors/researchers also highlight the benefits of a combined, integrative risk communication consisting of categories and probability scores (e.g., de Vogel et al., 2020). According to the current recommendations for risk assessment reports (Boetticher et al., 2019), the risk of recidivism posed by an assessed person must also be sufficiently specified: "(...) in particular, the probability of recidivism and type of offense must be substantiated" (p. 559).

Different international studies (Heilbrun et al., 2016; Singh et al., 2014) suggested a clear preference of forensic experts for a categorical risk communication over a dichotomous

or probabilistic form of communication. A survey study of risk assessment practice in Germany (Rettenberger at al., 2017) also showed that the preferred form of risk communication referred to a categorical assessment (in terms of the SPJ approach of "low," "moderate," or "high" risk).

6.1.6 Aim of the Study

In summary, the research results available from outpatient aftercare and socio-therapeutic institutions as well as the general practice of experts suggest a heterogeneous criminal risk assessment practice in Germany. However, since all data were based on surveys and interviews with assessors, the actual use of standardized risk assessment instruments - especially in everyday assessment practice - appears to be largely unclear. At present, it is also not known what the use of actuarial and clinical-structured prognostic instruments depends on in practice and how risk communication is carried out in assessment reports in which standardized risk assessment instruments were used.

The aims of the present study therefore consisted in examining:

(1) how frequently actuarial and clinical-structured prognostic instruments are used in criminal risk assessment practice,

(2) whether there are differences in the use of actuarial and clinical structured risk assessment instruments depending on

   a. report-related characteristics (time of contribution, institutional context, profession of the expert), as well as

   b. examinee-related characteristics (index delinquency, diagnosis, imprisonment or accommodation), and

(3) how risk communication is carried out in such expert reports in which standardized risk assessments instruments are used.

6.2 Method

6.2.1 Sample

The analysis units of the present study were *n* = 232 randomly selected external risk assessment reports about inmates from the penitentiary in Freiburg and *n* = 373 reports about individuals convicted of sexual and/or violent offenses from the Department of Forensic Psychiatry of the Clinic and Polyclinic for Psychiatry and Psychotherapy of Ludwig-Maximilian University in Munich (N = 605). Reports that were only based on records without personal examination of individuals, as well as incompletely archived reports, were excluded. The total number of assessment reports were ordered by diverse judicial parties to the proceedings (e.g. local or district courts, or higher regional courts).

In the university department, a total of n = 23 different department-affiliated experts were involved in the process of conducting assessment reports, whereas in the penitentiary in Freiburg a total of n = 71 external experts from independent expert practices, forensic psychiatric hospitals, psychological institutes, or psychotherapeutic and psychiatric hospitals or care centers were consulted. While at the Department of Forensic Psychiatry reports were conducted almost exclusively by psychiatric experts (n = 372, 99.7%), in the penitentiary in Freiburg external psychologists (n = 76, 32.8%), psychiatrists (n = 141, 60.8%), and experts of both professions (n = 15, 6.4%) were asked to assess examinees. Of the total sample of expert reports from the university department, 345 (92.5%) assessments were provided interdisciplinary (by consulting an additional psychological test examination or an additional exploration by a sociologist). In 15.1% (n = 35) of the expert reports of the penitentiary in Freiburg an interdisciplinary procedure was documented.

6.2.2 Data Collection and Material

To answer the study questions, the first author systematically analyzed the criminal risk assessment reports by a standardized questionnaire. First, report-related information, such as the orderer, the prognostic question, the year of contribution, the expert's profession, and

the use of additional psychological test reports or social science explorations, was gathered systematically. Furthermore, examinee-related characteristics such as age, gender, imprisonment or accommodation, previous and index delinquency, and (psychiatric) diagnoses according to ICD-10 were identified. The index offenses were divided into violent and sexual offenses, property/fraud offenses, offenses by narcotics law, and other offenses (e.g., property damage, arson). The use of risk assessment instruments was initially gathered dichotomously ("yes" = use of a risk assessment instrument of a corresponding instrument generation, "no" = no use of a risk assessment instrument of any generation). A distinction was made between second- and third-generation actuarial-statistical instruments and structured clinical instruments in the sense of "Structured Professional Judgement" (SPJ)[14]. In the present study, the second-generation actuarial-statistical instruments included the Static-99, Static-99-R, Static-2002, VRAG, SORAG, and the *Risk of Recidivism in Sex Offenders* (RRS; Rehder & Suhling, 2006), whereas the *STABLE-2000/2007*, the *Sex Offender Need Assessment Rating* (SONAR; Hanson & Harris, 2002), and the LSI-R were considered as instruments of the third-generation group. The SPJ instruments included the *Integrated List of Risk Variables* (ILRV; Nedopil, 2005), the HCR-20, the PCL-R, the *Psychopathy Checklist: Screening Version* (PCL-SV), the SVR-20, the Dittmann List, the so-called *Rasch Criteria* (Rasch, 1986), the FOTRES, the *Registrant Risk Assessment Scale* (RRAS; Ferguson et al., 1998; used as an SPJ instrument), and the *Short-Term Assessment of Risk and Treatability* (START; Webster et al., 2004). In expert reports in which standardized risk assessments instruments were used, the communication was gathered systematically and categorized into categorical (for example as "very high risk," "high risk", "moderate risk," "low risk and "very low risk "), dichotomous ("likely or unlikely to (...) commit further offenses"), and

[14] The PCL-R, frequently used in risk assessment reports, is a diagnostic assessment instrument for the clinical construct of *psychopathy* which is a robust predictor for persistent delinquency. Although it is not a classic risk assessment instrument, in the present study it was considered as SPJ tool because of the methodological approach.

quantitative (i.e., probabilistic, e.g., "...probability of recidivism within five years of 25%...") types of communication of risk.

6.3 Results

6.3.1 Statistical Sample Description of Risk Assessment Reports

Table 6.1 presents the total sample of risk assessment reports depending on included institutions, the judicial order, and offender-related characteristics (accommodation or imprisonment, index delinquency, psychiatric diagnoses according to ICD-10). Except for a few assessment reports on the preconditions for a preventive detention according to article 66 of the German penal code, the judicial order mostly questioned the release or the granting of privileges of individuals of forensic psychiatric hospitals or penitentiaries. More than 40% of the examinees were imprisoned in penitentiaries, while just under two-thirds were accommodated in forensic psychiatric hospitals according to articles 63, 64, or to preventive detention according to article 66 of the German penal code. More than 80% of the assessment reports were about individuals convicted of sexual or violent offenses, while the other offenses were property and fraud offenses, offenses by narcotics law, and other offenses (e.g., property damage, arson). No mental disorder was diagnosed in a third of the examinees. Personality and sexual preference disorders accounted for a good third of the diagnoses made, followed by schizophrenia, schizotypal and delusional disorders (F20-29, 9.8%), mental retardation (F70-79, 8.6%), and mental and behavioral disorders caused by psychotropic substances (F10-19, 6.9 %). Further diagnoses made in individual cases could be assigned to the disorder groups of affective disorders, neurotic, stress and somatoform disorders, developmental disorders and behavioral and emotional disorders with onset in childhood and adolescence. While the risk assessment reports from the LMU Munich contained almost all diagnosis groups according to ICD-10 and only about a quarter of the examinees had no diagnosis, the number of non-mentally ill individuals in the penitentiary was higher as

expected (56.1 %). The diagnoses here were almost exclusively distributed among personality and sexual preference disorders (just under 80 % of all diagnosed disorders).

Table 6.1

Statistical Sample Description of Risk Assessment Reports depending on Institution

Variables	Forensic Psychiatry, Munich		Penitentiary, Freiburg		Total	
	n	*%*	*n*	*%*	*n*	*%*
Judicial Order						
Privileges/Release of PEN/FPH[a]	305	81.8	135	58.2	440	72.7
Preconditions for article 66[b]	9	2.4	12	5.2	21	3.5
Privileges/Release of article 66[b]	59	15.8	85	36.6	144	23.8
Accomodation/Imprisonment						
Penitentiary	97	26.0	147	63.4	244	40.3
Preventive Detention article 66[b]	59	15.8	85	36.6	144	23.8
Forensic Psychiatry articles 63, 64[b]	217	58.2	0	0.0	217	35.9
Index Offence						
Violent Offence	163	43.7	90	38.8	253	41.8
Sexual Offence	97	26.0	47	20.3	144	23.8
Violent/Sexual Offence	72	19.3	43	18.5	115	19.0
Property/Fraud Offence	15	4.0	28	12.1	43	7.2
Offence by narcotics law	10	2.7	9	3.9	19	3.1
Other Offence	16	4.3	15	6.5	31	5.1
Psychiatric Diagnoses (ICD-10)						
No Diagnoses	99	26.5	130	56.1	229	37.9
F00-F09	13	3.5	1	0.4	14	2.3
F10-F19	22	5.9	20	8.6	42	6.9
F20-F29	59	15.8	0	0.0	59	9.8
F30-F39	8	2.1	0	0.0	8	1.3
F40-F49	2	0.5	0	0.0	2	0.3
F50-F59	0	0.0	0	0.0	0	0.0
F60-F69	147	39.4	81	34.9	228	37.7
F70-F79	20	5.4	0	0.0	20	8.6
F80-F89	1	0.3	0	0.0	1	0.2
F90-F99	2	0.5	0	0.0	2	0.3

Notes. $N = 605$ (Forensic Psychiatry, Munich: $n = 373$; Penitentiary, Freiburg: $n = 232$).

[a] PEN (Penitentiary, Freiburg), FPH (Forensic Psychiatric Hospital)

[b] articles 20, 21, 63, 64 or 66 of the German penal code

6.3.2 Use of Actuarial and Clinical-Structured Instruments in Practice

Table 6.2 provides the frequency of use of risk assessment instruments based on the historic generation of tools. Standardized risk assessment instruments were used in 270 of 605 risk assessment reports (44.6 %), which means that more than half of the analysed assessment reports were based exclusively on clinical-intuitive, experience-related predictions. While second-generation (n = 69, 11.4 %) and third-generation (n = 15, 2.5 %) actuarial risk assessment instruments were used only sporadically, clinical-structured, i.e., SPJ instruments were used in 261 prognostic reports (43.1 %). Applied second-generation actuarial instruments included the Static-99, Static-99-R, Static-2002, VRAG, SORAG, and RRS. Third-generation prognostic instruments included the STABLE-2000/2007, the SONAR, and the LSI-R. For clinical-structured judgment, the ILRV, the HCR-20, the PCL-R, the PCL-SV, the SVR-20, the Dittmann list, the Rasch criteria, the FOTRES, the RRAS, and the START were used.

Table 6.2

Frequency of Use of Different Risk Assessment Instruments Depending on the Generation

Risk assessment instruments	Frequency of use of the type of instruments	
	n	*%*
2nd generation	69	11.4
Static-99/99-R/2002	59	9.8
RRS	7	1.2
SORAG	6	1.0
VRAG	5	0.8
3rd generation	15	2.5
LSI-R	6	1.0
STABLE-2000/2007	6	1.0
SONAR	3	0.5
SPJ instruments	261	43.1
ILRV	193	31.9
HCR-20	78	12.9
PCL-R/SV	103	17.0
SVR-20	29	7.2
Dittmann list	28	4.6
Rasch criteria	4	0.7
FOTRES	3	0.5
RRAS	2	0.3
START	2	0.3
Total	270	44.6

Notes. $N = 605$

Tables 6.3 and 6.4 present the number of risk assessment reports including the use of actuarial and clinical-structured risk assessment instruments based on report-related (time of contribution, institutional context, profession of the expert) and examinee-related characteristics (imprisonment or accommodation, index delinquency, psychiatric diagnoses). Differences were found over time and with regard to the institutional context and the profession of the experts. Thus, the use of risk assessment instruments increased significantly over time. Standardized risk assessment approaches were used significantly more frequently in the university department in Munich than by external experts of the penitentiary in Freiburg. Exclusively considering the sub-sample of external experts at the penitentiary, psychological experts used actuarial and SPJ assessment instruments significantly more often than psychiatric experts.

In addition to the characteristics related to the assessment report, there were also differences depending on the characteristics of the examinees regarding the index delinquency, the diagnoses according to ICD-10, and the imprisonment or accommodation situation of examinees. The use of actuarial and clinical-structured assessment instruments was significantly more frequent among individuals who were imprisoned, that had a psychiatric diagnosis, and that were assessed for conviction of violent or sexual offenses. Except for the differences between the professions of experts and the accommodation or imprisonment of examinees due to the lack of comparison groups in the included institutions, all results could also be found separately for the respective institutions. The risk communication was categorical (from "very high risk" to "very low risk ") or dichotomous ("likely or unlikely to (...) commit further offenses") in just under two thirds of the total sample, quantitative in the form of probabilistic statements in only about 5%, and a combination of categorical and probabilistic forms of communication in just under one third (see Tab. 5).

Table 6.3

Frequency of Use of Actuarial and Clinical-Structured Risk Assessment Instruments depending on Report-Related Characteristics (Year of Contribution, Institutional Context, Profession of experts)

Use of risk assessment instruments	2nd generation		3rd generation		SPJ instruments		Total	
	n	*%*	*N*	*%*	*n*	*%*	*n*	*%*
Years of contribution				$\chi^2(2) = 102.287$, $p < .001$, $V = .411$				
1999 to 2004	1	0.5	0	0.0	40	18.7	40	18.7
2005 to 2010	37	13.3	6	2.2	141	50.7	148	50.7
2011 to 2016	31	27.4	9	8.0	80	70.8	82	70.8
Institution				$\chi^2(1) = 103.682$, $p < .001$, $V = .414$				
JVA Freiburg	8	3.4	4	1.7	43	18.5	43	18.5
LMU Munich	61	16.4	11	2.9	218	58.4	227	58.4
Profession of experts[a]				$\chi^2(1) = 12.737$, $p < .001$, $V = .234$				
Psychiatry (n = 156)	1	0.6	0	0.0	19	12.2	19	12.2
Psychology (n = 76)	7	9.2	4	5.3	24	31.6	24	31.6

Notes. N = 605.

[a] Due to a lack of comparison groups between the professions at the university department in Munich, only risk assessment reports of the penitentiary in Freiburg were considered (n = 232)

Table 6.4

Frequency of Use of Actuarial and Clinical-Structured Risk Assessment Instruments depending on Examinee-Related Characteristics (Accommodation or Imprisonment, Index Delinquency, Psychiatric Diagnoses according to ICD-10)

Use of risk assessment instruments	2nd generation		3rd generation		SPJ instruments		Total	
	n	*%*	*N*	*%*	*n*	*%*	*n*	*%*
Accomodation/Imprisonment				$\chi^2(2) = 65.889$, $p < .001$, $V = .330$				
Penitentiary	18	7.4	2	0.8	74	30.3	74	30.3
Preventive Detention article 66[a]	16	11.1	5	3.5	50	34.7	52	34.7
Forensic Psychiatry articles 63, 64[a]	35	16.1	8	3.7	137	63.1	144	63.1
Index Offence				$\chi^2(3) = 17.570$, $p = .001$, $V = .170$				
Violent Offence	5	2.0	1	0.4	114	45.1	114	45.1
Sexual Offence	36	25.0	7	4.9	67	46.5	85	46.5
Violent/Sexual Offence	27	23.5	6	5.2	97	48.7	85	48.7
Other Offence	1	1.1	1	1.1	97	25.8	85	25.8
Psychiatric Diagnoses (ICD-10)				$\chi^2(2) = 41.459$, $p < .001$, $V = .262$				
No Diagnoses	19	8.3	3	1.3	68	29.7	68	29.7
F60-F69	40	17.5	12	5.3	101	44.3	109	44.3
Other diagnoses	10	6.8	0	0.00	92	62.2	93	62.2

Notes. $N = 605$.

[a] articles 20, 21, 63, 64 or 66 of the German penal code

Table 6.5

Frequency of Use of Actuarial and Clinical-Structured Risk Assessment Instruments depending on Risk Communication

Type of risk communication	2nd generation (*n* = 69)		3rd generation (*n* = 15)		SPJ Instruments (*n* = 261)		Total (*n* = 270)	
	n	*%*	*N*	*%*	*n*	*%*	*n*	*%*
Categorical[a] / Dichotomous[b]	34	49.3	7	46.7	171	65.5	175	64.8
Quantitative[c]	35	2.2	0	0.0	13	5	14	5.2
Both	36	47.3	8	53.8	77	29	81	30.0

Notes. N = 270.

[a] Categorical = "very high risk," "high risk", "moderate risk," "low risk" and "very low risk"

[b] Dichotomous = "likely or unlikely to (...) commit further offenses"

[c] Quantitative = probabilistic, e.g., "...probability of recidivism within five years of 25%..."

6.4 Discussion

The results showed that actuarial or clinical-structured risk assessment instruments were only used in less than half of the total sample or assessment reports. Despite the impressive evidence in the forensic research literature of the higher predictive accuracy of standardized risk assessment compared to clinical-unstructured, intuitive, or predominantly or exclusively experience-based approaches (see, e.g., Rettenberger, 2018, 2019; Viljoen et al., 2021), a comparatively high proportion of so-called "clinical" (i.e., unstructured, intuitive, and experience-based) predictions was still evident in the present sample. The results were in line with previous empirical analyses of risk assessment reports (e.g., Haubner-Mclean & Eher, 2014) and supported previous survey results on criminal risk assessment practice in Germany (Rettenberger at al., 2017), which also indicated a use of standardized risk assessment instruments in approximately half of the risk assessments.

According to surveys (Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019), the standardization of criminal risk assessment seems to be more advanced in forensic institutions (social therapy, aftercare) than in external assessment practice. This could also be confirmed in the present study. The difference was found between the expert members of the university department and the external experts of the penitentiary. Again in line with previous studies (Rettenberger et al., 2017), results could also empirically support that the group of psychological experts made significantly more frequent use of risk assessment instruments than their psychiatric colleagues. This difference is certainly partly due to the different education focuses. The significant differences over time suggest a notable increase in the use of risk assessment instruments in expert witness practice and are consistent with comparable survey results in social therapy and aftercare risk assessment practice (Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019; Rettenberger et al., 2017).

In the present sample, mainly clinically structured, so-called SPJ instruments were used in the assessment reports, while actuarial instruments (of the 2nd and 3rd generation)

continued to be used relatively rarely across included institutions. Regardless of generation, the most frequently used assessment instruments largely coincided with previous survey results on the usefulness and application of assessment instruments in the German area (Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019; Rettenberger et al., 2017). Thus, it can be stated that in German-speaking risk assessment practice, several internationally established risk assessment instruments whose predictive validity is considered to be empirically validated, are applied regularly. However, there still seems to be a need for optimization, especially with regard to the use of actuarial instruments, which are empirically particularly well-validated (e.g. Hanson & Morton-Bourgon, 2009; Rettenberger, Rice, Harris & Eher, 2017).

Furthermore, the results of the present study show that the use of actuarial and clinical-structured risk assessment instruments appears to be partly dependent on examinee-related characteristics. It was shown that risk assessment instruments were used significantly more often in reports about individuals convicted of violent and/or sexual offenses. Comparable results were found for the presence of a psychiatric diagnosis according to ICD-10. Both findings could suggest that the severity of the index offense as well as the presence of a psychiatric diagnosis increases the probability of using standardized risk assessment instruments. This may be because a particularly large number of instruments have been developed for these offense groups, which could be empirically validated to the greatest extent. Additionally, these groups of offenses and diagnoses may pose special challenges for experts, which they try to overcome by relying on empirically validated methods.

The examination of risk communication in the present sample of assessment reports revealed a clear preference for a categorical form of communication, which is also in line with international studies (Heilbrun et al., 2016; Singh et al., 2014) and those of the German-speaking area (Rettenberger at al., 2017). In just under two-thirds of the total sample, risk was communicated categorically or dichotomously. However, in just under one-third - particularly

in the sub-sample of the university department - a combined form of communication could be identified, as it has been increasingly proposed or called for by authors in the recent past (including de Vogel et al., 2020; Eher et al., 2019; Hanson et al., 2017a; 2017b). The predominantly categorical risk communication could also be attributed to the more frequent use of clinical-structured risk assessment instruments compared to actuarial tools in the present sample, as the nominal form of communication is the method of choice for the former group (von Franqué, 2013).

When interpreting the results, it should generally be noted that the two sub-samples (assessments contributed by the university department and the external experts of the penitentiary) are only comparable to a limited extent and that the university institution certainly has a special status, as stronger institutional quality assurance can be assumed (Wertz et al., 2018). This was reflected not only in the use of risk assessment instruments, but also in the involvement of additional psychological test examinations and social science explorations as part of a generally more pronounced interdisciplinary approach at the university institution. While experts at the university department processed risk assessment reports on individuals of both forensic psychiatric hospitals and the penitentiary, the sub-sample of the penitentiary only contained risk assessment reports on prisoners and persons in preventive detention regarding article 66 of the German penal code. Consequently, not only did the judicial orders and the imprisonment or accommodation of the examinees differ, so did the prevalence of (psychiatric) diagnoses according to ICD-10. Nonetheless, the results regarding the use of risk assessment instruments were cross-institutional, from which generalizable conclusions can be drawn. It could therefore be concluded that a tendency towards standardized judgment based on risk assessment instruments in criminal prognostics is increasing not only in university assessment practice but in general.

Even though the use of scientifically based risk assessment instruments can add a significant contribution to the validity of criminal risk assessment reports and the added value

of standardized instruments for risk assessment has been emphasized in the current recommendations for risk assessment reports (Kröber et al., 2019), their use can not be absolutely required in every case, since there are also cases in which the use of standardized instruments would not make sense (e.g., in a very rare case or in crime constellations that were not or could not be considered in the development samples). In this respect, it cannot necessarily be said that there is a lack of quality if experts do not use standardized instruments. Generally applicable guidelines or recommendations for the use of special risk instruments cannot be formulated and always depend on the context and the specific prognostic question. In principle, the selection of the appropriate methodological approach for the risk assessment is the responsibility of the expert. Nevertheless, it already follows from the formulation of the methodological minimum requirements or recommendations that statistical methods should be used regularly (Kröber et al., 2019). Empirical research results also impressively underline the added value of standardized assessment methodology. Irrespective of the fact that German case law stipulates a high degree of individualization of risk assessment (Boetticher et al., 2019), standardized risk assessment instruments provide empirically sound statements that should be an essential decision support in the required individual case assessment. On the one hand, the present results speak for an increasing standardization of criminal risk assessments, while on the other hand show a further need for quality assurance of criminal risk assessment.

7. A Comparison of the Predictive Accuracy of Structured and Unstructured Risk Assessment Methods for the Prediction of Recidivism in Individuals Convicted of Sexual and Violent Offense[15]

Abstract

One of the most commonly replicated results in the research area of recidivism risk assessment is the superiority of structured and standardized prediction methods in comparison to unstructured, subjective, intuitive or impressionistic clinical judgments. However, the quality of evidence supporting this conclusion is partly still controversially discussed because studies including direct comparisons of the predictive accuracy of unstructured and structured risk assessment methods have been relatively rarely conducted. Therefore, in the present study we retrospectively examined $N = 416$ expert witness reports written about individuals convicted of violent and/or sexual offenses in Germany between 1999 and 2015. The predictive accuracy of different methodological approaches of risk assessment (subjective clinical [i.e., unstructured clinical judgment; UCJ], structured professional judgment [SPJ], actuarial risk assessment instruments [ARAIs], and combinations of ARAIs-/SPJ-based risk assessments) was compared by analysing the actual reoffenses according to the Federal Central Register (average follow-up period $M = 7.08$ years). In accordance with previously published results, the results indicated a higher predictive accuracy for structured compared to unstructured risk assessment approaches for the prediction of general, violent, and sexual recidivism. Taken together, the findings underline the limited accuracy of unstructured clinical judgments and provided further support for the use of structured and standardized risk assessment procedures in the area of crime and delinquency.

Keywords: Sexual offenses, violent offenses, clinical prediction, actuarial prediction,

[15] Paper published as Wertz, M., Schobel, S., Schiltz, K. & Rettenberger, M. (in press). A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. Psychological Assessment.

structured professional judgment.

7.1 Introduction

There are different methodological approaches to risk assessment of individuals convicted of violent and/or sexual offenses. These approaches could be basically divided into subjective clinical (or unstructured, intuitive, unguided, or impressionistic), actuarial (also including statistical, mechanical, or algorithmic) as well as structured professional or clinical-idiographic predictions (e.g., Grove et al., 2010; Meehl, 1954; Nicholls et al., 2013). Subjective clinical judgments are defined as assessments based solely on clinical experience and judgment of assessors using informal and subjective methods, which are predominantly justified by their training, expertise, and professional designations (Grove et al., 2010; Hanson & Morton-Bourgon, 2009; Skeem & Monahan, 2011). This approach is referred to as "unstructured" because of its lack of explicit rules for assessors, which increases its vulnerability to biases and as a consequence its limited reliability and validity (Bengtson & Långström, 2007; Nicholls et al., 2013). Contrarily, highly structured actuarial risk assessment instruments (ARAIs) contain a predetermined list of empirically derived static and/or dynamic risk factors and a statistical or algorithmic model to combine the risk factor scores into a total score (Grove et al., 2010; Nicholls et al., 2013). Structured professional judgment (SPJ) approaches are professional guidelines integrating also the risk-related information derived from a predetermined list of risk (and sometimes protective) factors. However, contrary to ARAIs, the application of SPJ guidelines is based on an idiographic interpretation of the relevance of each factor and implies a clinical (rather than an algorithmic or statistical) processing of the risk-related data (Nicholls et al., 2013). In a nutshell, a decision based on ARAIs is usually made mechanically according to a fixed algorithm, while in SPJ instruments the final decision is made by a structured assessment based on personal

expertise and the theoretical and empirical knowledge of a professionally educated assessor (Lodewijks et al., 2008).

7.1.1 Generations of Risk Assessment Approaches

In terms of generations (Andrews et al., 2006; Taxman, 2017), unstructured clinical judgment [UCJ] is considered as the first generation of risk assessment, followed by the second (consisting of static risk factors) and third generations (containing dynamic risk factors) of ARAIs and the SPJ approaches (fourth generation). The clinical-idiographic method combines actuarial knowledge and structured professional judgments with theoretically sound explanations of the individual behaviors including clinical (e.g., neuropsychological or psychopathological) aspects of the examinee by strictly following scientific standards (Craig & Rettenberger, 2018; Craig et al., 2020).

7.1.2 Superiority of Structured Methods

When it comes to research about the comparison between these different assessment approaches, several studies have underlined the superiority of structured methods (second, third, and fourth generation) and the limited accuracy of UCJ (first generation) not only for criminal risk assessment settings, but for a diverse range of different aspects of human behaviors (e.g., Ægisdóttir et al., 2006; Bengtson & Långström, 2007; Hanson & Morton-Bourgon, 2009). A number of previously published studies consistently highlighted, that unstructured assessments were significantly more susceptible to biases (e.g., Grove et al., 2000; Johansen, 2006; Turgut et al., 2006). Particularly for the prediction of sexual or violent recidivism, the structured method is considered as clearly superior compared to unstructured approaches (e.g., Bonta et al., 1998; Heilbrun et al., 2010; Jackson et al., 2004). Therefore, in the last decades a multitude of ARAIs and SPJ tools for assessing the recidivism risk of different populations and for different settings and outcomes has been developed (e.g., Guy, 2008; Singh et al., 2014; Viljoen et al, 2021). However, when direct comparisons between ARAIs and SPJ instruments have been carried out no clear and consistent superiority for

either has been identified (e.g., Campbell et al., 2009; Douglas et al., 2005; Hanson et al., 2007).

7.1.3 Ongoing Debate on Limited Accuracy of Unstructured Clinical Judgments

Given this frequently reported superiority of structured risk assessment approaches compared to unstructured and intuitively made judgments and its consequent widespread use in forensic institutions and in expert witness reports (e.g., Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019; Singh et al., 2014), it seems to be at the first glance surprising that various aspects in this field of assessment research and practice are still discussed controversially. In this sense, several studies and reviews pointed out the methodological limitations of the existing research (e.g., Dressel & Farid, 2018; Litwack, 2001; Mossman, 1994) and indicated that the accuracy of unstructured assessments might not be as limited as stated (e.g., de Vogel et al., 2004; Lin et al., 2020; Melton et al., 2018). For example, a recent umbrella review of systematic reviews comparing structured and unstructured risk assessment methods reported that "although research is generally consistent in reporting that risk assessment tools are superior to UCJ, studies used to support this statement showed serious problems in terms of risk for bias and lack of direct comparison" (Viljoen et al., 2021, p. 92). The authors showed, for example, that almost the entire state of research was conducted decades ago and nearly two thirds of the primary studies included in most reviews were published in the 1980s or earlier. Furthermore, only a few studies compared directly structured and unstructured assessment approaches and did not examine whether the predictive validity differed significantly but based their conclusions on a visual inspection of the results. Several studies did not focus particularly on risk assessment instruments and did not provide direct head-to-head comparisons between ARAIs or SPJ tools and UCJs. Given these research desiderata, further empirical studies which directly compare unstructured with structured risk assessment methods are highly recommended (Viljoen et al., 2021).

7.1.4 Study Objectives

The present study tries to remedy this research need by directly comparing the predictive accuracy of unstructured as well as different structured risk assessment approaches derived from a comprehensive sample of German risk assessment reports about individuals convicted of violent and/or sexual offenses. More precisely, the first study objectives were to identify and classify the applied assessment approaches in the risk assessment reports due to their degree of structuring the assessment process. Second, we aimed to compare the predictive accuracy of unstructured and structured risk assessment approaches (i.e., unstructured clinical judgment [UCJ], ARAIs [in terms of the revised version of the Violence Risk Appraisal Guide, VRAG-R; Rice et al., 2013], SPJ instruments, and combinations of ARAIs-/SPJ-based risk assessments). In order to examine the stability and generalizability of the results, different follow-up periods (5-year and in total) and offence types (general, nonviolent, violent, general sexual, and sexual contact recidivism) have been used. Based on the existing state of research we hypothesized that structured risk assessment methods would outperform unstructured risk assessment procedures in predicting general, sexual (contact) and violent recidivism. However, between the different types of structured risk assessment methods we did not expect significant differences with the exception that we would expect the highest predictive accuracy for the combination of different structured assessment approaches (i.e., the combined use of SPJ and ARAIs).

7.2 Method

In order to collect the data for present study, $N = 416$ expert witness reports about individuals convicted of violent and/or sexual offenses have been retrospectively evaluated to gather information about the current and previous delinquency, the psychiatric diagnosis (according to ICD-10), and the time of incarceration or placement in forensic psychiatric units

due to articles 20, 21, 63, 64 or 66 of the German penal code[16]. Furthermore, various data related to the expert witness reports (timeframe, institutional context, expert profession, use of risk assessment tools, methodological approach) were collected. The VRAG-R ratings were collected retrospectively by officially trained psychologists by using the German version of the instrument (Rettenberger et al., 2017). Finally, the predictive accuracy of the different assessment approaches was compared by actual recidivism data derived from the Federal Central Criminal Register of the Federal Office of Justice. This study was approved by the institutional review board of the University Regensburg, Germany (protocol number 17-805-101); the study was not preregistered.

7.2.1 Sample

The sample consisted of $N = 416$ risk assessment reports about male individuals charged or convicted of violent and/or sexual offenses, which have been gathered in two German institutions representing common forensic practice: the penitentiary in Freiburg ($n = 184$), and the Department of Forensic Psychiatry of the University Hospital Munich ($n = 232$)[17]. Given the aim to compare unstructured and structured risk assessments on the same group of individuals, we excluded persons who have been convicted of nonviolent or nonsexual offenses, females, and persons aged 17 or younger and those who were not yet released during the follow-up period, or lacking a criminal record because of death, emigration, or unclear identification. Reports that were only based on records without

[16] In Germany, defendants, who are not responsible and who are at the same time considered as dangerous, are sentenced to mandatory treatment in a forensic mental hospital. Those who are diminished responsible are sentenced to both hospital treatment and prison. Individuals suffering from addiction disorders are sentenced by article 64 of the German penal code and from any other diagnoses by article 63 of the German penal code. Furthermore, Germany also has provisions for preventive detention of persons regarded as dangerous (section 66 of the German penal code). Risk assessments will be required from mental health professionals not only at the time of sentencing but several times thereafter because detaining and releasing mentally ill or chronically dangerous individuals from hospitals or prisons depend strongly on the results of their risk assessment reports (Dunn et al., 2014).

[17] The University Hospital Munich provides external forensic assessment reports for a diverse number of different courts and public prosecutors. The penitentiary in Freiburg is run by the federal state Baden-Wuerttemberg in the southwest of Germany and pursues the primary goal of the rehabilitation of the prisoners. At a penitentiary, assessment reports are not done by the institution itself but are instead regularly obtained by external psychiatric or psychological experts who are in most cases not affiliated to a specific academic or scientific institution.

personal examination of the individuals as well as incompletely archived reports were also excluded. Assessment reports conducted between 1999 and 2015 were ordered by diverse judicial parties of the proceedings like local or district courts, courts responsible for the execution of the sentence, higher regional courts, and public prosecutors. In general, the risk assessment reports included in the present study were written by 86 different experts. More than three quarters of assessments were contributed primarily by psychiatrists (84.6%, *n* = 352), while 15.4% (*n* = 64) were reported by psychologists. Two hundred forty-seven (59.4%) criminal risk reports were conducted and reported by interdisciplinary teams (e.g., in primary responsibility by a psychiatrist but supported by an additional psychological test report or a sociological interview about the social rehabilitation process)[18].

7.2.2 Measures

Risk Communication. In order to categorize the final risk judgment of each report, a five-point Likert-scale was used ("very low risk", "low risk", "moderate risk", "high risk", and "very high risk") in accordance to current risk communication guidelines recommending five levels of recidivism risk, as they provide a framework for standardizing risk communication by matching the information contained in risk tools to a risk classification system which is independent of any particular risk scale widely applicable (Hanson et al., 2017). To additionally examine the interrater reliability, a randomly selected and independently coded sample of 208 final judgments was co-rated, which resulted in an excellent reliability coefficient of the coding scores (ICC = .866, 95%-CI =.801 to .914).

Unstructured Clinical Judgment (UCJ). We defined unstructured clinical judgments (UCJs) as assessments in which risk factors were selected, measured, and combined solely based on subjective clinical experience and intuition (e.g., Grove et al., 2010; Hanson &

[18] All experts of the university hospital were members of the Department of Forensic Psychiatry and therefore extensively experienced in forensic evaluations as well as supervised by the head of department; however,for the external psychiatric or psychological experts, which have written the reports for the penitentiary, no additional data (e.g., years of experience) were available.

Morton-Bourgon, 2009; Skeem & Monahan, 2011). If an ARAI or SPJ tool was used, assessments were considered as *structured*. Afterwards, the assessment process was further categorized regarding the type (ARAIs, SPJ, or both) and generation (second, third and fourth generation or combinations).

Second and Third Generation of Risk Assessment Tools. Second generation ARAIs used in the present study were the German versions of the *Static-99* (Hanson & Thornton, 2000; Rettenberger et al., 2013), *Static-2002* (Hanson & Thornton, 2003), *Sex Offender Risk Appraisal Guide* (SORAG; Quinsey et al., 2006; Rettenberger et al., 2017), and the *Violence Risk Appraisal Guide* (VRAG; Quinsey et al., 2006), whereas the German versions of the *Level of Service Inventory-Revised* (LSI-R; Andrews & Bonta, 1995; Dahle et al., 2012), *Stable-2007* (Hanson & Harris, 2007; Matthes & Rettenberger, 2008), and *Sex Offender Need Assessment Rating* (SONAR; Hanson & Harris, 2002) were considered as third generation tools.

Structured Professional Judgment (SPJ). SPJ instruments were represented by the German versions of the *Integrative List of Risk Variables* (ILRV; Nedopil, 2005; Stübner et al., 2006), the *Historical Clinical Risk Management-20* (HCR-20; Müller-Isberner et al., 1998; Webster et al., 1997), and *Sexual Violence Risk-20* (SVR-20; Boer et al., 1997; Müller-Isberner et al., 2000). The *Psychopathy Checklist-Revised* (PCL-R; Hare, 2003; Mokros et al., 2017) is a structured diagnostic instrument for diagnosing the degree of psychopathic personality traits, which is a robust predictor for persistent delinquency (Rice et al., 2013; Singh et al., 2011; Yang et al., 2010). The German version of the PCL-R (Mokros et al., 2017) obtained also adequate indices for its reliability of the test scores and (predictive) validity of the test score interpretations (e.g., Mokros et al., 2014). Although the PCL-R was originally not developed as a risk assessment tool, it is regularly used in the international and German-speaking risk assessment practice (Singh et al., 2014). As PCL-R scores are

integrated into various SPJ instruments like, for example, the HCR-20 and the SVR-20, the PCL-R was allocated to the SPJ category in the present study.

VRAG-R. Finally, the German version of the VRAG-R (Rettenberger at al., 2017; Rice et al., 2013) was rated retrospectively for a random subsample (n = 230) due to resource complaints. The ratings were made by two authors (MW, SS) of the present study who have been officially trained in the use of the VRAG-R. All ratings were conducted blind to actual criminal recidivism data. The subsample consisted of 107 individuals convicted of sexual, and 123 convicted of violent offenses. Reports that did not provide sufficient information to ensure an adequate rating of the VRAG-R items were excluded. There were no significant differences regarding the sample characteristics (i.e., no significant differences regarding the proportion of incarceration vs. forensic psychiatric placement, the institutional place, the distribution of mental disorders, and the number of prior offenses) between this subsample and the total sample. The VRAG-R[19] consists of twelve predominantly static items, the total score could range from -34 to 46 and could be divided into nine risk bins (Rettenberger et al., 2017; Rice et al., 2013). The German version of the VRAG-R yielded also high effect sizes for its reliability of the test scores and (predictive) validity of the test score interpretations which were comparable to the original version (Gregório Hertz et al., 2021a, 2021b). In the present study, the interrater reliability of the VRAG-R was calculated by analysing the agreement of the test scores between the independent ratings made by two authors (MW, SS) in a subsample of n = 23 randomly selected assessment reports (10.0%) resulting in an excellent reliability coefficient (ICC = .934, 95%-CI = .924 to .949).

Recidivism Data. To directly compare the predictive accuracy of UCJs, ARAIs, SPJ instruments as well as risk assessment judgments based on a combined use of ARAIs and SPJ

[19] Whereas the VRAG was developed in terms of assessing violent recidivism in general, the derivative version called Sex Offender Risk Appraisal Guide (SORAG) is as a tool for assessing the risk of a violent recidivism in individuals who have committed sexual offenses. As stated in the revision study (Rice et al., 2013), the VRAG-R replaces both the VRAG and the SORAG by providing an instrument for assessing violent recidivism in any individual who committed at least one offense (Gregório Hertz et al., 2021a; Rice et al., 2013).

instruments, recidivism data of the participants were coded from the official criminal records derived from the Federal Central Criminal Register of the Federal Office of Justice in June 2016. For the present study, two different timeframes were used: first, complete follow-up periods of each participant (average follow-up period M = 7.08 years) were analysed, following by analyses which used the fixed 5-years follow-up periods. Recidivism was coded as general (each new conviction of any kind), nonviolent (each new conviction because of a nonviolent offense), sexual (each new conviction involving both sexual contact and noncontact offenses), sexual contact (only new sexual offenses including physical contact[20]), or violent (each new conviction because of violent offenses) recidivism. At all ratings before, both raters did not have any information about actual criminal records. All data were analysed using IBM, SPSS, Version 26.0. Due to privacy, legal, and ethical restrictions of the raw data of the present study, it was not possible to make the data publicly available; however, it is possible to receive access to the data and the material of the present data by sending a request to the first author.

7.2.3 Data Analysis

We examined the ability of the different risk assessment approaches to discriminate between recidivists and nonrecidivists by calculating the *Area Under the Curve* (AUC) derived from *Receiver Operating Characteristics* (ROC). AUC values represent the probability that a randomly selected recidivist has a higher score on a measure than a nonrecidivist (Helmus & Babchishin, 2017). They are an indicator of the global performance of a predictor, not being sensitive to base rate effects and allowing direct comparisons of different scales and could be regarded as the most commonly used and recommended statistic for risk assessment scales (Rice & Harris, 2005). Referring to Rice and Harris (2005), AUC values were interpreted as follows: AUC ≥ .72 were classified as “large”, AUC = .64 to .71 as

[20] All individuals convicted of a sexual offense including violence were coded as “sexual contact offense”.

"moderate", and statistically significant AUC ≤ .63 as "small". Differences between AUC values were tested for significance by using a *z*-statistic (DeLong et al., 1988). Furthermore, we calculated *Odds Ratios* (ORs) derived from logistic regressions (e.g., Helmus & Babchishin, 2017) by using again fixed follow-up time periods of five years for a subsample of $n = 312$ individuals. Therefore, reoffenses after five years have not been considered for this kind of analysis. The ORs provide information about the increase of the recidivism risk between two randomly selected adjacent total scores. To conduct a time-independent examination of the predictive accuracy, we calculated *Rate Ratios* (RRs) derived from Cox regression analyses which control for unequal follow-up periods allowing to consider the total sample consisting of participants with varying follow-up periods (Allison, 1984). RRs are a measure for the relationship between the probabilities rather than the odds of two groups.

7.3 Results

Table 7.1 shows the descriptive statistics of the characteristics of the total sample and separately for the subsamples of reports about individuals charged or convicted of sexual, sexual contact, and violent offenses. More than half of the reports were about individuals diagnosed with a mental disorder according to ICD-10, who were placed in a preventive detention or forensic psychiatry. Only a small number had not been previously convicted for at least one offense before the index offense. In terms of the final judgments of the reports, more than half of the examinees were described to have a (very) low or a moderate recidivism risk, while the remainder of the reports yielded a high or even a very high risk. Considering the methodological approach, at least one actuarial or SPJ risk assessment tool was used in nearly half of the reports, which are therefore considered as structured. On the contrary, the other reports were defined as UCJs (see Table 1).

Table 7.1

Offender- and Report-Related Characteristics among Risk Assessments about Individuals Convicted of Violent and/or Sexual Offenses (N = 416)

Characteristic	Sexual offenses (*n* =172)		Sexual contact offenses (*n* = 125)		Violent offenses (*n* = 244)		Total sample (*n* =416)	
	n	%	*n*	%	*n*	%	*n*	%
Offender-related								
Incarceration/Placement								
Penitentiary	72	41.9	52	41.6	121	49.6	193	46.4
Preventive Detention §66[a]	55	32.0	49	39.2	47	19.3	102	24.5
Forensic Psychiatry §63/64[a]	45	26.2	24	19.2	76	31.1	121	29.1
ICD-10 Diagnoses[b]								
No Diagnosis	71	41.3	51	40.8	105	43.0	176	42.3
F10-F19	11	6.4	9	7.2	20	8.2	31	7.5
F20-F29	5	2.9	3	2.4	32	13.1	37	8.9
F60-F69	79	45.9	60	48.0	70	28.7	149	35.8
Other	6	3.5	2	1.6	17	7.0	23	5.5
Prior offenses								
None	11	6.4	8	6.4	9	3.7	20	4.8
Any	161	93.6	117	93.6	235	96.3	396	95.2
Nonviolent	14	8.1	12	9.6	38	15.5	52	12.5
Violent	15	8.7	12	9.6	178	73.0	193	46.4
Sexual	34	19.8	13	10.4	0	0.0	34	8.2
Sexual contact	98	57.0	80	64.0	19	7.8	117	28.1
Report-related								
Timeframe								
1999 to 2004	64	37.2	48	38.4	86	35.2	150	36.1
2005 to 2010	82	47.7	59	47.2	120	49.2	202	48.5
2011 to 2015	26	15.1	18	14.4	38	15.6	64	15.4

Characteristic	Sexual offenses (n = 172)		Sexual contact offenses (n = 125)		Violent offenses (n = 244)		Total sample (n = 416)	
	n	%	*n*	%	*n*	%	*n*	%
Institution								
University hospital	93	54.1	56	44.8	139	57.0	232	55.8
Penitentiary	79	45.9	69	55.2	105	43.0	184	44.2
Risk assessment tools								
Actuarial tools	44	22.5	28	22.4	6	2.5	50	12.0
2. Generation[c]	38	22.1	26	20.8	5	2.0	43	10.3
3. Generation[d]	6	3.5	4	3.2	1	0.4	7	1.7
SPJ tools[e]	83	48.3	62	49.6	100	40.9	183	44.0
Methodological approach								
Unstructured clinical (UCJ)	89	51.7	63	50.4	144	59.0	233	56.0
Structured (SPJ)	45	26.2	36	28.8	94	38.5	139	33.4
Structured (SPJ + actuarial)	38	22.1	26	28.8	6	2.5	44	10.6
Assessment judgment								
Very low risk	16	9.3	8	6.4	26	10.7	42	10.1
Low risk	64	37.2	44	35.2	104	42.5	168	40.4
Moderate risk	13	7.6	10	8.0	20	8.2	33	7.9
High risk	46	26.7	35	28.0	68	27.9	114	27.4
Very high risk	33	19.2	28	22.4	26	10.7	59	14.2

Note. Frequencies of occurrence are shown for the index offense categories sexual offenses, sexual contact offenses, and violent offenses. Sexual offenses were defined as crimes against sexual self-determination including sexual harassment/coercion, sexual assault, rape, sexual abuse, purchase, possession and distribution of materials containing child/ youth pornography, exhibitionism, procuring, and forced prostitution. Sexual contact offenses were defined as sexual offenses including physical contact. Violence offense was defined as nonsexual violence including intentional killing, murder, manslaughter, assault, robbery, hostage taking, blackmail, coercion, deprivation of liberty/abduction, and arson.

[a] Articles 20, 21, 63, 64 or 66 of the German penal code

[b] According to the International Statistical Classification of Diseases and Related Health Problems (ICD-10): F00-F09 = Organic, including symptomatic, mental disorders; F10-F19 = Mental and behavioural disorders due to psychoactive substance use; F20-F29 = Schizophrenia, schizotypal and delusional disorders; F30-F39 = Mood [affective] disorders; F40-F49 = Neurotic, stress-related and somatoform disorders; F50-F059 = Behavioural syndromes associated with physiological disturbances and physical factors; F60-F69 = Disorders of adult personality and behaviour; F70-F79 = Mental retardation; F80-F89 = Disorders of psychological development; F90-F99 = Behavioural and emotional disorders with onset usually occurring in childhood and adolescence

[c] Represented in the sample by the German versions of the Static-99, Static-2002, SORAG, and VRAG

[d] Represented in the sample by the German versions of the LSI-R, Stable-2007, and SONAR

[e] Represented in the sample by the German versions of the ILRV, HCR-20, SVR-20, and PCL-R

7.3.1 Recidivism Rates

The sample was followed-up for 7.08 years on average (*SD* = 4.14, range from 1.5 to 17 years). The average time between the submission of the assessment report and the release of the examinees was 1.85 years (*SD* = 2.18, range from 0.5 and 3 years). Table 7.2 presents the recidivism rates of the total sample and the subsample with a fixed 5-year follow-up period. Of the total sample, almost half were reconvicted for a general re-offense during the complete follow-up period. Individuals convicted of sexual offenses showed significantly lower recidivism rates than those convicted of violent offenses for general, $\chi^2(1) = 5.379$, $p < .05$, $\varphi = .114$, as well as for violent recidivism, $\chi^2(1) = 18.948$, $p < .001$, $\varphi = .213$. No significant differences could be found regarding sexual, $\chi^2(1) = 3.144$, $p = .076$, $\varphi = .087$, sexual contact, $\chi^2(1) = 1.505$, $p = .220$, $\varphi = .060$, and nonviolent, $\chi^2(1) = .142$, $p = .706$, $\varphi = .018$ recidivism. The subsample with the fixed 5-year follow-up periods showed necessarily lower recidivism rates due to less time of exposure to risk of recidivism of examinees (see Table 2 for details).

Table 7.2

Recidivism Rates among Individuals Convicted of Violent and/or Sexual Offenses for Total Follow-Up (N = 416) and 5-Year at-Risk (n = 312)

Recidivism	Sexual offenses (*n* = 172)		Sexual contact offenses (*n* = 125)		Violent offenses (*n* = 244)		Total sample (*n* = 416)	
	n	%	*n*	%	*n*	%	*n*	%
Total follow-up[d]								
None	103	59.9	74	59.2	118	48.4	221	53.1
Any[a]	69	40.1	51	40.8	126	51.6	195	46.9
Nonviolent recidivism	53	30.8	39	31.2	71	29.1	124	29.8
Violent recidivism[b]	7	4.1	5	4.0	50	20.5	57	13.7
Sexual recidivism[c]	9	5.2	7	5.6	5	2.0	14	3.4
Sexual contact recidivism	4	2.3	4	3.2	4	1.6	8	1.9
Five-year follow-up								
None	82	62.6	62	61.4	91	50.3	173	55.4
Any[a]	49	37.4	39	38.6	90	49.7	139	44.6
Nonviolent recidivism	36	27.5	28	27.7	52	28.7	88	28.2
Violent recidivism[b]	7	5.3	5	5.0	33	18.2	40	12.8
Sexual recidivism[c]	6	4.6	6	5.9	5	2.8	11	3.5
Sexual contact recidivism	3	2.3	3	3.0	4	2.2	7	2.2

Note. Frequencies of occurrence are shown for the index offense categories sexual offenses, sexual contact offenses and violent offenses. Sexual offenses were defined as crimes against sexual self-determination including sexual harassment/coercion, sexual assault, rape, sexual abuse, purchase, possession and distribution of materials containing child/ youth pornography, exhibitionism, procuring, and forced prostitution. Sexual contact offenses were defined as sexual offenses including physical contact. Violence offense was defined as nonsexual violence including intentional killing, murder, manslaughter, assault, robbery, hostage taking, blackmail, coercion, deprivation of liberty/abduction, and arson. Recidivism was coded from criminal records according to the Federal Central Register.

[a] Each new criminal conviction of any kind

[b] Nonsexual violence regarding intentional killing, murder, manslaughter, assault, robbery, hostage taking, blackmail, coercion, deprivation of liberty/abduction, and arson

[c] Crimes against sexual self-determination including sexual harassment/coercion, sexual assault, rape, sexual abuse, purchase, possession and distribution of materials containing child/ youth pornography, exhibitionism, procuring, and forced prostitution

[d] Average time at-risk (calculated from date retrieved from criminal records) was within the index category sexual offenses 6.93 years (*SD* = 3.92), within the index category sexual contact offenses 7.03 years (*SD* = 3.9) and within the index category violent offenses 7.19 years (*SD* = 4.29) as well as within the total sample 7.08 years (*SD* = 4.14)

7.3.2 Predictive Validity Using Variable Follow-Up Periods

Table 7.3 gives an overview about the predictive validity of the different methodological approaches analysed in the present study (UCJ, ARAIs in terms of VRAG-R scores, SPJ, and the combination of SPJ and ARAIs) by using two different effect sizes (AUC and RR) and variable follow-up periods; the latter allowed the inclusion of the total sample (N = 416). As can be seen in Table 3, five different recidivism categories (any, nonviolent, violent, sexual, and sexual contact recidivism) as well as different subsamples (sexual, sexual contact, and violent offenses) were used in order to examine the predictive validity of the different methodological approaches. In accordance with our hypothesis that UCJ provided only limited predictive performance, UCJ showed only weak (or no significant at all) predictive validity indices for the prediction of any, nonviolent, violent, sexual, and sexual contact recidivism. Based on the existing state of research we hypothesized that structured risk assessment methods would generally outperform UCJ judgements, which was generally speaking also confirmed by the results of the present study. However, this general conclusion has some important constraints, which could be plausibly interpreted against the background of the methodological approaches. The pure actuarial approach (i.e., the retrospectively collected VRAG-R data) yielded a particularly high predictive accuracy for the prediction of violent recidivism, which is the outcome measure the instrument was originally designed for. This finding could be interpreted as an indicator for the outcome specificity of the VRAG-R in particular and the actuarial prediction approach in general (Rettenberger et al., 2017).

Similarly, the SPJ approach showed also a higher predictive accuracy than the UCJ-based predictions. However, as expected the highest predictive validity was found for the combination of SPJ methods and ARAIs. Even for the prediction of recidivism in the subsamples of sexual and sexual contact offenses (which was a comparatively difficult task given the relatively small sample sizes and low recidivism rates), the combined use of structured risk assessment methods provided extraordinary high effect sizes. As can be seen in

Table 7.3, a general limitation of the analysis refers to the fact that for some recidivism criteria and subsamples the analysis of the predictive accuracy was not possible due to low recidivism base rates and too small sample sizes. In the next step, we analysed whether these differences in the predictive validity reached the threshold of statistical significance. For the total sample, effect sizes were found to differ significantly between the methodological prediction approaches for general, violent, and sexual recidivism (UCJs vs. SPJ + ARAIs: $z = 2.367$, $p < .05$ for general recidivism and $z = 5.672$, $p < .001$ for violent recidivism; SPJ vs. SPJ + ARAIs: $z = 2.980$, $p < .01$ for violent recidivism; UCJ vs. SPJ: $z = 2.225$, $p < .05$ for sexual recidivism). Considering only individuals convicted of violent offenses, significant differences were found for general (UCJ vs. SPJ: $z = 2.215$, $p < .05$) and nonviolent recidivism (UCJ vs. SPJ: $z = 2.211$, $p < .05$). In both subsamples of individuals convicted of sexual and sexual contact offenses, effect sizes were found to differ significantly for violent recidivism (UCJ vs. SPJ + ARAIs: $z = 3.319$, $p < .001$ in individuals convicted of sexual offenses and $z = 2.270$, $p < .05$ in those convicted of sexual contact offenses). Effect sizes for individuals convicted of sexual contact offenses also showed significant differences between UCJ and the combined SPJ and ARAIs assessments in predicting general recidivism ($z = 1.980$, $p < .05$). For all comparisons of dependent AUC estimates, the differences of effect sizes between the VRAG-R total scores and UCJ indicated also a higher predictive accuracy for ARAIs but its superiority failed to reach statistical significance for general ($z = 1.612$, $p = .081$) and violent recidivism ($z = 1.742$, $p = .071$).

Table 7.3

AUC Values and Cox Regression Analyses (Rate Ratios; RR) of the Different Methodological Approaches and Actuarial Total Scores (Variable Total Follow-Up) (N = 416)

Recidivism	Sexual offenses (*n* = 172)				Sexual contact offenses (*n* = 125)				Violent offenses (*n* = 244)				Total sample (*N* = 416)			
	AUC		RR		AUC		RR		AUC		RR		AUC		RR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Any																
Unstructured clinical (UCJ)	.56^{ns}	[.44, .68]	1.41*	[1.13, 86]	.51^{ns}	[.36, .65]	1.36^{ns}	[0.97, 1.89]	.62*	[.53, .72]	1.45*	[1.22, 1.73]	.59*	[.51, .63]	1.43*	[1.24, 1.65]
Actuarial (VRAG-R)	.65*	[.54, .77]	1.03**	[1.01, .05]	.62^{ns}	[.47, .77]	1.03*	[1.01, 1.06]	.68**	[.59, .78]	1.02**	[1.01, 1.04]	.66**	[.59, .73]	1.03**	[1.02, 1.04]
Structured (SPJ)	.57^{ns}	[.40, .74]	1.47**	[1.04, 1.98]	.61^{ns}	[.42, .79]	1.38^{ns}	[1.00, 1.94]	.77**	[.67, .87]	1.78**	[1.33, 2.36]	.69**	[.60, .78]	1.53**	[1.08, 1.75]
Structured (SPJ + actuarial)	.73*	[.54, .92]	2.17**	[1.23, .86]	.75*	[.54, .96]	1.97*	[1.02, 3.79]	.87^{ns}	[.55, .99]	1.74^{ns}	[0.97, 3.57]	.79**	[.64, .94]	1.81**	[1.20, 2.43]
Nonviolent																
Unstructured clinical (UCJ)	.56^{ns}	[.43, .68]	1.47^{ns}	[1.11, 1.96]	.51^{ns}	[.36, .66]	1.34^{ns}	[0.92, 1.94]	.58^{ns}	[.48, .68]	1.49^{ns}	[1.16, 1.92]	.57^{ns}	[.50, .65]	1.47^{ns}	[1.01, 1.63]
Actuarial (VRAG-R)	.64*	[.52, .77]	1.03**	[1.01, 1.06]	.56^{ns}	[.40, .74]	1.02^{ns}	[1.00, 1.05]	.52^{ns}	[.41, .63]	1.01^{ns}	[0.99, 1.03]	.59^{ns}	[.49, .66]	1.02**	[1.01, 1.04]
Structured (SPJ)	.58^{ns}	[.41, .77]	1.48^{ns}	[0.73, 1.98]	.51^{ns}	[.31, .71]	1.34^{ns}	[1.04, 1.99]	.74**	[.63, .84]	1.82**	[1.30, 2.54]	.65**	[.55, .74]	1.67**	[1.23, 1.79]
Structured (SPJ + actuarial)	.60^{ns}	[.37, .83]	1.55^{ns}	[0.79, 3.05]	.57^{ns}	[.32, .82]	1.41^{ns}	[0.52, 2.87]	.71^{ns}	[.22, .99]	1.51^{ns}	[0.78, 3.67]	.67^{ns}	[.48, .86]	1.53^{ns}	[0.96, 2.20]
Violent																
Unstructured clinical (UCJ)	.58^{ns}	[.37, .70]	1.24^{ns}	[0.65, 2.38]	.54^{ns}	[.28, .71]	1.18^{ns}	[0.50, 2.78]	.56^{ns}	[.45, .67]	1.42^{ns}	[1.11, 1.80]	.52^{ns}	[.42, .62]	1.29^{ns}	[1.06, 1.66]
Actuarial (VRAG-R)	.71^{ns}	[.49, .92]	1.05^{ns}	[0.99, 1.11]	.71^{ns}	[.40, .99]	1.05^{ns}	[0.97, 1.13]	.73**	[.63, .84]	1.04**	[1.02, 1.07]	.71**	[.61, .80]	1.05**	[1.02, 1.07]
Structured (SPJ)	.73^{ns}	[.48, .98]	1.85^{ns}	[0.31, 9.10]	.74^{ns}	[.49, .99]	1.92^{ns}	[0.46, 9.68]	.64^{ns}	[.46, .81]	1.47^{ns}	[0.85, 2.54]	.67^{ns}	[.48, .77]	1.67^{ns}	[0.82, 2.10]
Structured (SPJ + actuarial)	.91*	[.82, .99]	2.98^{ns}	[0.47, 9.51]	.91*	[.80, .99]	2.68^{ns}	[0.49, 9.46]	.90^{ns}	[.62, .99]	---	---	.89*	[.79, .99]	2.73*	[1.15, 6.51]
Sexual																
Unstructured clinical (UCJ)	.52^{ns}	[.28, .75]	1.32^{ns}	[0.72, 2.43]	.57^{ns}	[.34, .70]	1.21^{ns}	[0.56, 2.58]	.69^{ns}	[.45, .94]	2.23^{ns}	[0.97, 5.14]	.59^{ns}	[.41, .78]	1.77^{ns}	[1.07, 2.95]
Actuarial (VRAG-R)	.61^{ns}	[.38, .84]	1.03^{ns}	[0.98, 1.07]	.77^{ns}	[.56, .99]	1.10^{ns}	[0.97, 1.25]	.82^{ns}	[.76, .89]	1.09^{ns}	[0.96, 1.23]	.66^{ns}	[.47, .84]	1.04^{ns}	[0.99, 1.08]
Structured (SPJ)	.73^{ns}	[.48, .98]	2.57^{ns}	[0.57, 9.41]	.74^{ns}	[.49, .99]	2.81^{ns}	[0.37, 9.08]	.96^{ns}	[.90, .99]	6.86^{ns}	[0.71, 9.78]	.86*	[.72, .99]	5.86*	[0.71, 9.78]
Structured (SPJ + actuarial)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Recidivism	Sexual offenses (n = 172)				Sexual contact offenses (n = 125)				Violent offenses (n = 244)				Total sample (n = 416)			
	AUC		RR		AUC		RR		AUC		RR		AUC		RR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sexual contact																
Unstructured clinical (UCJ)	.65^{ns}	[.53, .78]	3.42^{ns}	[0.70, 9.08]	.59^{ns}	[.45, .73]	3.17^{ns}	[0.64, 9.65]	.85^{ns}	[.67, .99]	5.06^{ns}	[0.77, 9.08]	.75^{ns}	[.64, .87]	4.16^{ns}	[1.39, 9.88]
Actuarial (VRAG-R)	.80*	[.62, .97]	1.08^{ns}	[0.98, 1.20]	.78^{ns}	[.61, .94]	1.09^{ns}	[0.95, 1.26]	.81^{ns}	[.74, .88]	1.07^{ns}	[0.92, 1.25]	.79*	[.69, .90]	1.08*	[0.99, 1.18]
Structured (SPJ)	.73^{ns}	[.48, .98]	4.32^{ns}	[0.73, 9.18]	.74^{ns}	[.49, .99]	3.81^{ns}	[0.68, 9.14]	.96^{ns}	[.90, .99]	6.41^{ns}	[0.72, 9.98]	.86*	[.72, .94]	5.88*	[0.72, 9.98]
Structured (SPJ + actuarial)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Note. Sexual offenses were defined as crimes against sexual self-determination including sexual harassment/coercion, sexual assault, rape, sexual abuse, purchase, possession and distribution of materials containing child/ youth pornography, exhibitionism, procuring, and forced prostitution. Sexual contact offenses were defined as sexual offenses including physical contact. Violence offense was defined as nonsexual violence including intentional killing, murder, manslaughter, assault, robbery, hostage taking, blackmail, coercion, deprivation of liberty/abduction, and arson. Recidivism was coded from criminal records according to the Federal Central Register. If a cell is not filled, data were not applicable due to sample size.

CI = confidence interval; ns = nonsignificant

*p < .05. **p < .01.

7.3.3 Predictive Validity Using Fixed Follow-Up Periods

Table 7.4 shows the AUC values and the ORs of the different risk assessment approaches among a subsample with a fixed follow-up time of five years. For UCJ both methods for calculating effect sizes (AUC and ORs) yielded a similar result pattern compared to the above-mentioned variable follow-up periods: For the total sample as well as for all subgroups the UCJ provided no or only small (for predicting any recidivism in the subsample of violent offenses and in the total sample) predictive accuracy, whereas the actuarial approach (represented by the VRAG-R total scores) yielded significant moderate to large effect sizes for the prediction of general (any), violent, sexual, and sexual contact recidivism only. Similarly, risk assessments based on SPJ instruments yielded also significant moderate to large effect sizes for different types of recidivism but again only for the total sample and the subsample of violent offenses. Finally, the risk assessments based on both SPJ and ARAIs showed extraordinarily large significant effect sizes for the prediction of violent (also for the comparatively small subsample of sexual offenses), sexual, and sexual contact recidivism (the latter at least for the total sample).

Table 7.4

AUC Values and Logistic Regression Analyses (Odds Ratio; OR) of the Different Methodological Approaches and Actuarial Total Scores (Fixed Follow-Up of Five Years) (N = 312)

Recidivism	Sexual offenses (*n* =131)				Sexual contact offenses (*n* = 101)				Violent offenses (*n* = 181)				Total sample (*N* = 312)			
	AUC		OR		AUC		OR		AUC		OR		AUC		OR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Any																
Unstructured clinical (UCJ)	.61^{ns}	[.48, .74]	1.39^{ns}	[0.97, 1.98]	.55^{ns}	[.40, .71]	1.21^{ns}	[0.79, 1.84]	.58*	[.48, .68]	1.28*	[0.96, 1.69]	.58*	[.51, .67]	1.31*	[1.03, 1.59]
Actuarial (VRAG-R)	.64*	[.51, .77]	1.03*	[1.01, 1.07]	.61^{ns}	[.44, .77]	1.03^{ns}	[0.99, 1.07]	.78**	[.68, .88]	1.07**	[1.03, 1.10]	.72**	[.63, .80]	1.05**	[1.03, 1.07]
Structured (SPJ)	.65^{ns}	[.46, .84]	1.47^{ns}	[0.86, 2.49]	.71^{ns}	[.51, .92]	1.70^{ns}	[0.92, 3.13]	.86**	[.76, .97]	3.86**	[1.97, 7.56]	.75**	[.64, .85]	2.15*	[1.41, 2.94]
Structured (SPJ + actuarial)	.65^{ns}	[.36, .95]	1.78^{ns}	[0.79, 3.99]	.71^{ns}	[.40, .99]	2.03^{ns}	[0.81, 5.07]	---	---	---	---	.73^{ns}	[.49, .98]	2.03*	[1.04, 4.46]
Nonviolent																
Unstructured clinical (UCJ)	.61^{ns}	[.49, .75]	1.18^{ns}	[0.75, 2.02]	.57^{ns}	[.41, .73]	1.24^{ns}	[0.79, 1.94]	.51^{ns}	[.41, .61]	1.01^{ns}	[0.74, 1.37]	.55^{ns}	[.47, .64]	1.15^{ns}	[0.91, 1.46]
Actuarial (VRAG-R)	.57*	[.43, .72]	1.02^{ns}	[0.99, 1.06]	.55^{ns}	[.36, .73]	1.02^{ns}	[0.98, 1.07]	.53^{ns}	[.39, .66]	1.01^{ns}	[0.98, 1.04]	.55^{ns}	[.46, .65]	1.01^{ns}	[0.99, 1.03]
Structured (SPJ)	.65^{ns}	[.46, .78]	1.21^{ns}	[0.72, 2.04]	.62^{ns}	[.38, .85]	1.34^{ns}	[0.74, 2.40]	.77**	[.64, .90]	2.32**	[1.34, 3.99]	.67**	[.55, .79]	1.60**	[1.12, 2.27]
Structured (SPJ + actuarial)	.66^{ns}	[.37, .83]	1.36^{ns}	[0.84, 2.77]	---	---	---	---	---	---	---	---	.66^{ns}	[.48, .81]	1.55^{ns}	[1.05, 2.19]
Violent																
Unstructured clinical (UCJ)	.61^{ns}	[.37, .76]	1.08^{ns}	[0.45, 2.57]	.57^{ns}	[.34, .79]	1.42^{ns}	[0.34, 8.25]	.57^{ns}	[.44, .70]	1.26^{ns}	[0.91, 1.76]	.58^{ns}	[.42, .66]	1.13^{ns}	[0.85, 1.52]
Actuarial (VRAG-R)	.71^{ns}	[.45, .87]	1.02^{ns}	[0.96, 1.09]	---	---	---	---	.86**	[.76, .95]	1.12**	[1.05, 1.19]	.79**	[.69, .89]	1.08**	[1.04, 1.13]
Structured (SPJ)	.86^{ns}	[.67, .99]	2.19^{ns}	[0.91, 6.35]	.88^{ns}	[.70, .99]	2.18^{ns}	[0.81, 6.85]	.80*	[.68, .92]	2.59*	[0.91, 7.35]	.81*	[.62, .87]	2.24*	[0.97, 5.51]
Structured (SPJ + actuarial)	.94*	[.82, .99]	5.25*	[1.10, 9.78]	.92*	[.78, .99]	5.01*	[1.02, 9.98]	---	---	---	---	.90*	[.76, .99]	3.55*	[1.17, 9.83]
Sexual																
Unstructured clinical (UCJ)	.52^{ns}	[.29, .76]	1.12^{ns}	[0.57, 2.24]	.55^{ns}	[.33, .68]	1.04^{ns}	[0.47, 1.91]	.70^{ns}	[.46, .94]	1.81^{ns}	[0.77, 4.27]	.61^{ns}	[.44, .80]	1.43^{ns}	[0.84, 2.44]
Actuarial (VRAG-R)	.79^{ns}	[.62, .95]	1.08^{ns}	[0.97, 1.20]	.76^{ns}	[.57, .96]	1.08^{ns}	[0.96, 1.24]	.80^{ns}	[.72, .89]	1.08^{ns}	[0.96, 1.22]	.79*	[.70, .88]	1.08*	[1.00, 1.17]
Structured (SPJ)	.86^{ns}	[.67, .99]	3.04^{ns}	[1.96, 6.31]	.88^{ns}	[.70, .99]	3.24^{ns}	[1.76, 6.51]	.96^{ns}	[.90, .99]	4.65^{ns}	[1.98, 7.71]	.92*	[.84, .99]	3.34^{ns}	[1.97, 6.51]
Structured (SPJ + actuarial)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Recidivism	Sexual offenses (n = 131)				Sexual contact offenses (n = 101)				Violent offenses (n = 181)				Total sample (n = 312)			
	AUC		OR		AUC		OR		AUC		OR		AUC		OR	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
Sexual contact																
Unstructured clinical (UCJ)	.65^{ns}	[.51, .79]	1.70^{ns}	[0.60, 4.79]	.57^{ns}	[.41, .72]	1.43^{ns}	[0.51, 4.04]	.85^{ns}	[.68, .99]	4.33^{ns}	[0.57, 9.66]	.75^{ns}	[.64, .87]	2.41^{ns}	[0.95, 6.09]
Actuarial (VRAG-R)	.78^{ns}	[.53, .99]	1.07^{ns}	[0.95, 1.22]	.75^{ns}	[.47, .99]	1.08^{ns}	[0.93, 1.25]	.78^{ns}	[.69, .88]	1.07^{ns}	[0.91, 1.25]	.77*	[.65, .90]	1.07^{ns}	[0.97, 1.18]
Structured (SPJ)	.86^{ns}	[.67, .99]	3.16^{ns}	[1.79, 6.21]	.88^{ns}	[.70, .99]	3.36^{ns}	[1.79, 6.93]	.96^{ns}	[.90, .99]	6.65^{ns}	[1.98, 9.71]	.92*	[.84, .94]	3.56^{ns}	[1.99, 6.81]
Structured (SPJ + actuarial)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Note. Sexual offenses were defined as crimes against sexual self-determination including sexual harassment/coercion, sexual assault, rape, sexual abuse, purchase, possession and distribution of materials containing child/ youth pornography, exhibitionism, procuring, and forced prostitution. Sexual contact offenses were defined as sexual offenses including physical contact. Violence offense was defined as nonsexual violence including intentional killing, murder, manslaughter, assault, robbery, hostage taking, blackmail, coercion, deprivation of liberty/abduction, and arson. Recidivism was coded from criminal records according to the Federal Central Register. If a cell is not filled, data were not applicable due to sample size.

CI = confidence interval; ns = nonsignificant

*p < .05. **p < .01.

7.4 Discussion

Despite the widespread use of structured risk assessment approaches (e.g., Singh et al., 2014; Tully et al., 2013; Viljoen et al, 2021), there seems to be still a relatively heterogeneous application practice of risk assessment methods: More than half of the reports examined in the present study were considered as UCJs which underlines the ongoing controversial debate on the relevance and benefits of intuitively and subjectively made assessment judgments (e.g., Dressel & Farid, 2018; Lin et al., 2020; Melton et al., 2018).

7.4.1 Superiority of Structured Compared to Unstructured Risk Assessment Approaches

The results of the present study provide an empirical comparison of the predictive accuracy of unstructured and structured methodological approaches of risk assessment in a sample of German risk assessment reports. In accordance with previously published results and meta-analytic findings (e.g., Ægisdóttir et al., 2006; Bengtson & Långström, 2007; Turgut et al., 2006), the comparisons presented here indicate a higher predictive accuracy for structured compared to unstructured risk assessment approaches for the prediction of general, violent, and sexual recidivism across different subgroups. When using the total follow-up time (average follow-up of 7.08 years) as well as when using the fixed 5-year follow-up time periods, both methods for calculating effect sizes (AUC and ORs/RRs) yielded a comparable result pattern in terms of a limited predictive accuracy of UCJs. For all comparisons, the differences of effect sizes between structured and unstructured risk assessment methods indicated a greater predictive accuracy of ARAIs and SPJ instruments compared to unstructured methods, although not all comparisons reached of the threshold of statistical significance. The latter finding might be at least partly due to the well-known fact of comparatively low recidivism base rates for some subsamples (particularly for the subsample of individuals convicted of sexual offenses; Rettenberger et al., 2015). Thereby, no clearly distinct superiority for either actuarial or SPJ approaches was identified; this finding

confirmed similar results reported in previous studies which compared both methods (e.g., French, 2021; Hanson & Morton-Bourgon, 2009; Nicholls et al., 2013). From a statistical point of view it is noteworthy that the reported result pattern remained virtually unchanged when we considered different effect sizes (AUC, ORs, or RRs). These findings support also previous research (e.g., Etzler et al., 2020; Helmus, & Babchishin, 2017; Rettenberger et al., 2013) and provide further support for the stability of AUC values as an index for the predictive accuracy of risk assessment instruments even in smaller samples (Hanczar et al., 2010).

7.4.2 Outcome Specificity of Structured Risk Assessment

With regard to the results of the predictive accuracy of the different structured risk assessment approaches, different conclusions could be drawn. First, the results of the present study replicate previously published data which have shown that the test scores of the German version of the VRAG-R yielded comparable predictive validity indices as the original version (Gregório Hertz et al., 2021a, 2021b). The VRAG-R was significantly predictive for (sexual) violent and general reoffenses, which can be interpreted as further evidence for its predictive validity across different countries, jurisdictions, and language areas. Given that the present study was conducted by researchers who were not part of the (Anglo-American) development team of the VRAG-R, the present findings can be also interpreted as evidence against the previously discussed allegiance effect and instead in favor of a fidelity effect (Blair et al., 2008; Harris et al., 2010). Furthermore, as can be for the VRAG-R the present results provided generally additional support for the outcome specificity of structured risk assessment approaches given that the instruments tended to perform better for the outcome criterion for which they were originally designed for (Hanson & Morton-Bourgon, 2009; Rettenberger et al., 2017). On the other hand, the different risk assessment approaches yielded generally speaking lower predictive accuracy indices for a comparatively broad outcome criterion like general recidivism or a relatively narrowly defined outcome criterion like sexual (contact)

recidivism. From a diagnostic point of view this finding is not surprising given the fact that the most commonly used ARAIs and SPJ instruments in Germany are designed to predict violent (including sexual) recidivism (like the VRAG-R or the HCR-20; Rettenberger et al., 2017).

7.4.3 Combination of ARAIs and SPJ Instruments

A further noteworthy result refers to the finding that the combination of ARAIs and SPJ instruments yielded particularly high effect sizes for the prediction of (general, violent, and sexual) recidivism, which were predominantly higher than the predictive accuracy of ARAIs or SPJ instruments alone. This finding supports the idea of a convergent risk assessment approach where the strengths of both ARAIs and SPJ instruments are combined (Singer et al., 2016). Consequently, the current generation of SPJ instruments like, for example, the recent third version of the HCR-20 (HCR-20^{V3}; Douglas et al., 2014) recommends also the consideration of the results of ARAIs in the course of the formulation of the final risk judgment. For applied risk assessment settings one obvious suggestion for integrating both approaches into one final risk judgment could be that if both ARAIs and SPJ instruments indicated a comparable risk assessment judgment (i.e., both the VRAG-R as well as the HCR-20^{V3} indicated a relatively low—or moderate or high—risk), then the final (combined, convergent) judgment is clear. However, if both risk assessment approaches led to different conclusions (i.e., the VRAG-R indicated a high risk, whereas the application of the HCR-20^{V3} indicated a low or moderate risk), then the crucial challenge of the assessor is to unravel this assessment discrepancy (i.e., by explaining why the more clinically and dynamically oriented assessment of the HCR-20^{V3} yielded a lower risk than the more static and statistical approach of the VRAG-R).

7.4.4 Practical Implications

Taken together, from an applied risk assessment perspective, the findings of the present study indicate, first, that assessors should use indeed both risk assessment approaches

and should combine and integrate the results of both into the final judgment. Second, the competition of ARAIs and SPJ representatives in the past seems to be resolved in the present tense: The crucial issue is not which approach is better but rather how to combine both in order to achieve the best prediction. The combination of both approaches is also relevant from another point of view: The use of ARAIs alone has inevitable limitations because the actuarial approach is not able to provide an individual (idiographic) explanatory model of delinquency. This means that a clinical-idiographic approach is necessary and even mandatory in the German penal law which is also confirmed by currently published recommendations and methodological minimum requirements for criminal risk assessment (Kröber et al., 2019). It is important to emphasize that a comprehensive and individually based risk assessment approach is therefore indispensable and the pure use of group-based statistics or nomothetic rules is insufficient (Ægisdóttir et al., 2006). However, individual clinical and case-specific risk assessment of examinees should though not be conducted intuitively but should always follow strictly scientific, transparent, and evidence-based standards. Therefore, current comprehensive risk assessment models have tried to integrate the individual and relatively flexible approach of case formulation with the standardized assessment approach of ARAIs and SPJ instruments (e.g., the *Case Formulation Incorporating Risk Assessment* [CAFIRA] model; Craig & Rettenberger, 2018; Craig et al., 2020).

7.4.5 Limitations

A general limitation of the present study which needs to be addressed is the retrospective research design. In this context, it is not possible to ascertain whether assessors using validated, empirically supported risk assessment instruments are maybe in general more experienced or trained in conducting forensic evaluations compared to experts who do not use these methods. Confounding variables of assessors themselves may further explain differences in accuracy of risk assessments, so it may be a promising approach to control for different assessor characteristics in futures studies. Methodologically, experimental designs

including a random assignment to a specific approach and the use of prospective-longitudinal field study designs would be the best way for (cross-)validating risk assessment approaches and instruments (e.g., Nitsche et al., 2022; Rettenberger et al., 2017). However, particularly experimental research approaches can be accompanied by serious ethical conflicts and legal issues. Given the fact that one main aim of the present study was to identify clinically applied risk assessment methods in their everyday practice, the retrospective design was indispensable. Another data-related limitation was that the information about recidivism was only based on one single data source (i.e., officially registered reconvictions in accordance to the Federal Central Register), which is inevitably an underestimation of the actual degree of repeated delinquency. A further relevant aspect for the interpretation of the present findings is the fact that most assessment reports were conducted by clinicians employed by a university hospital. Therefore, it is not entirely clear to what extent the results of the present study could be generalized to clinicians and expert witnesses who were not part of an academic institution.

It is also important to note that we defined the degree of structure of the assessment reports by considering if and what underlying type and generation of risk assessment instrument (ARAIs vs. SPJ instruments) was used, i.e., if an ARAI or SPJ tool was used, assessments were considered as structured. Of course, this procedure could be interpreted as only a rough approximation to a sophisticated categorization of risk assessment approaches. Furthermore, there were standardised assessment procedures like the PCL-R, which can only hardly be allocated to either of the before mentioned groups of risk assessment methods. Finally, in order to be able to contrast a “pure” actuarial risk assessment approach with other assessment methods, it was necessary to score the VRAG-R retrospectively and to compare these retrospective assessment results with the prospectively made assessments of the expert witness reports. However, as previously published studies about second generation ARAIs indicated (e.g., Nitsche et al., 2022; Rettenberger et al., 2013, 2017), the predictive accuracy

of this actuarial approach yielded comparable results independently of the applied research design (i.e., if a retrospective of a prospective data collection procedures was used).

7.4.6 Conclusion

In accordance with a number of previously published results (e.g., Grove et al., 2010; Nicholls et al., 2013; Skeem & Monahan, 2011), the comparisons between different risk assessment approaches in the present study indicate a higher predictive accuracy for structured compared to unstructured risk assessment approaches for the prediction of different outcome criteria (general, violent, and sexual recidivism). Given the weight of evidence in support of structured risk assessment, the findings underline the limited accuracy of UCJs and provide further support for the use of structured and standardized risk assessment procedures in the field of forensic sciences and psychological assessment of crime, delinquency, and recidivism. However, a comprehensive and convergent risk assessment approach including clinical-idiographic evaluation of the individual biography and background is not only mandatory with regard to the German penal law but yielded also the highest effect sizes for predictive accuracy (compared to ARAIs and SPJ instruments alone). As different previous studies and reviews indicated that the predictive accuracy of unstructured assessments could be possibly better as originally stated given the variety of methodological limitations of the existing research data (e.g., Lin et al., 2020; Melton et al., 2018; Viljoen et al., 2021), the present study remedied this research gap by overcoming at least some of the discussed methodological limitations of prior research. But even if the present findings basically underline the importance and the relatively high predictive validity of structured risk assessment approaches, the need for further research became also clear, particularly concerning direct comparisons of the predictive accuracy of different methodological approaches in applied risk assessment settings.

8. General Discussion and Conclusion[21]

To date, research literature of the last few decades points to a relatively heterogeneous quality of expert witness reports (Nowara, 1995a, 1995b; Suhling, 2003; Dahle et al., 2009, 2012; Haarig et al., 2012; Kunzl & Pfäfflin, 2011; see Chapter 1.2.1). Since the publication of methodological minimum requirements for risk assessment reports in 2006 in Germany and updated recommendations in 2019, there is little empirical evidence on whether and how these standards are put into clinical and judicial practice (see Chapter 1.2.1.1). It also remains unclear, how frequently psychometric tests as well as actuarial and clinical-structured prognostic instruments are used in real criminal assessment practice in Germany (see Chapter 1.2.1.3). Given the frequently reported superiority of structured risk assessment approaches compared to unstructured and intuitively made judgments, various aspects of standardization in this field of assessment research and practice are still discussed controversially (see Chapter 1.2.1.2). Therefore, the overall aim of this thesis was to systematically examine the quality of criminal risk and responsibility assessment reports in German practice (see Chapter 1.3).

[21] Parts of the Discussion are published in Wertz, M., Kury, H. & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [The application of methodological minimum requirements for risk assessment reports in clinical practice - An empirical validation using officially registered reoffenses]. Forensische Psychiatrie, Psychologie, Kriminologie, 12(1), 51 – 60; Wertz, M., Schiltz, K., Imhoff, R.& Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosebegutachtung [The influence of the judicial order on the quality of the work of expert witnesses in the context of risk assessment]. Recht & Psychiatrie, 38(4), 193 – 200; Wertz, M., Hausam, J., Konrad, N., Schiltz, K., Imhoff, R.& Rettenberger, M. (2021). Qualität von Schuldfähigkeitsgutachten – Mindestanforderungen, unterbringungsrelevante Gefährlichkeitsprognose und Berücksichtigung im richterlichen Urteil [Quality of criminal responsibility reports – Minimum requirements, risk assessment, and consideration in court decisions]. Recht & Psychiatrie, 39(4), 202 – 211; Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R. & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. Psychology, Crime & Law. Advance online publication. https://doi.org/10.1080/1068316X.2022.2063286; Wertz, M., & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale [The use of standardized risk assessment instruments in the practice of risk assessment: Empirical findings on frequency and risk communication as a function of assessment- and subject-related characteristics.]. Forensische Psychiatrie und Psychotherapie, 28(3), 241-261, and in Wertz, M., Schobel, S., Schiltz, K. & Rettenberger, M. (in press). A comparison of the predictive accuracy of structured and unstructured risk assessment methods for the prediction of recidivism in individuals convicted of sexual and violent offense. Psychological Assessment.

Concluding, the present thesis offers several strengths regarding existing research gaps and methodological issues. First, an extensive sample of more than 1.000 risk and criminal responsibility reports from different institutions was used, representing common general practice in Germany. Second, a diverse population of individuals convicted or charged of sexual and/or violent offenses of penitentiaries or psychiatric hospitals were included, also representing common report- (time of contribution, institutional context, profession of the expert, judicial order, methodological approach) and examinee-related (sociodemographic data, index and prior delinquency, imprisonment, or accommodation mental disorders) characteristics in general practice. Third, due to the retrospective study design it was possible to consider officially registered re-offenses for different follow-up periods. Therefore, criminal records were derived from the Federal Central Criminal Register of the Federal Office of Justice. Fourth, also judicial verdicts (or corresponding judicial documents available) were additionally gathered for a subsample of reports, so information on the outcome of the respective proceedings from the ordering public prosecutors' offices and courts was obtained. Fifth, quality of assessment reports was systematically operationalised by application of methodological minimum requirements, published by an interdisciplinary working group of leading clinicians, researchers, and expert witnesses in this field. Therefore, a survey instrument based on the methodological minimum requirements (Boetticher et al., 2006) and the relevant research literature (Dahle, 2005a, 2005b; Nedopil, 2005; Kröber, 2006) was developed in accordance with previous studies (Kury & Adams, 2010; Riegl, 2007). Sixth, while at least some studies are available for the international, primarily Anglo-American area (e.g., Archer et al., 2006; Singh et al., 2014; Viljoen et al., 2010), only a few empirical studies have so far been conducted for the German-speaking area on the form in which criminal risk assessments are made in professional practice. Since all data was based on surveys, interviews with assessors, or self-reports of clinicians (e.g., Archer et al., 2006; Lally, 2003; Wright et al., 2017), the actual use of standardized risk assessment instruments

and psychometric tests - especially in everyday assessment practice - appears to be largely unclear. Thus, the present thesis is the first investigation to provide systematically recorded data about the real-world use of actuarial and structured-professional risk assessment instruments and psychometric tests in expert witness reports in real actual German practice. Finally, to validate the quality of all retrospective ratings (e.g., implementation of minimum requirements, prognostic judgements, recidivism, hit rates), the reliability criterion was met by calculating the interrater reliability with experienced, in-depth and intensively trained co-rater of psychological profession and scientific expertise. It was constantly determined on a randomly selected and independently coded sample of expert reports, resulting in reliability coefficients that can be classified as very high according to the usual standard (Leonhart, 2004).

In summary, as most scientific references pointing to a heterogenous quality are based on subjective impressions of experienced expert witnesses or presentations of individual cases without profound empirically sound foundation, the present thesis is the first to provide an empirically sound extensive examination of the quality of expert witness reports in real German practice.

8.1 Contributions to Empirical Scientific Evidence

Thus, the present thesis expands research knowledge in several ways, presented as follows: Regarding the implication of **methodological minimum requirements for risk and criminal responsibility reports** (Boetticher et al., 2006; 2007), the comparison of the assessment reports before and after the publication of the methodological minimum requirements shows that there has been a significant improvement in quality over time (see Chapter 2). This indicates an improvement in the quality of expert assessment practice as a result of the publication of the methodological minimum requirements. The developed methodological minimum requirements seem to have arrived in practice, but there is still an enormous potential for improvement concerning the practical application of the criteria,

reflecting the lasting notable heterogeneity of quality. Even if the methodological minimum requirements have gained general appreciation in the specialised sciences in the meantime, they are by no means taken into account by every forensic expert (Dahle et al., 2012; Kunzl et al., 2009; Kunzl & Pfäfflin, 2011; Schnoor, 2009; Verell, 2015). It can therefore for the future be assumed that research recommendations, contrary to the efforts of the initiators, will not necessarily lead to the desired effect across institutions and in a timely manner.

With regard to the differences in risk assessment reports concerning the institutional comparison in favour of the university hospitals, the research proximity of a university department can be discussed, since university institutions have a special status in the assessment report practice. This conclusion is also supported by the observed differences in quality depending on the institutional affiliation of the external experts at the penitentiary, underlining that research proximity seems so be more decisive than profession of experts. The occurrence of inaccurate risk assessments can be reduced, but not eliminated by the compliance with the minimum requirements for risk assessment reports (Dahle, 2005b; 2006). However, in the present thesis it could be shown that compliance with the methodological minimum requirements is associated with a higher accuracy of the risk assessments. It can be concluded that experts should be encouraged more consistently to consider and apply the minimum requirements, as these clearly contribute to an increase in the quality of the risk assessment reports and thus to a more valid prognostic assessment of the recidivism risk of the examinees.

In this respect, it seems obvious that research findings from this area should also be accounted for to a greater extent in the risk assessment for possible placement to mandatory treatment in criminal responsibility assessment reports, as also supported by the recently updated recommendations for risk assessment reports (Boetticher et al., 2019; Kroeber et al., 2019). The present examination of risk assessment methods in criminal responsibility reports in which the orderer questioned the prerequisites of a mandatory treatment revealed a

heterogeneous implementation from a quantitative and qualitative point of view (see Chapter 4). Only half of the reports in which a risk assessment would have been indicated contained a risk assessment, which was predominantly unstructured. The scope of the assessment in the written report varied greatly, showing no significant changes over time. Consequently, scientific findings in risk assessment research do not seem to have arrived in criminal responsibility assessment practice.

The methodological minimum requirements were primarily intended for forensic experts, but also for judges, public prosecutors and defense attorneys, facilitating criminal risk assessment reports and their evaluation of their validity. Thus, the minimum requirements were also to assist in delimiting which questions were to be the subject of the expert assessment and which questions would have to be answered exclusively by the court alone. The results of the present thesis suggest a stronger orientation of the judicial orders towards the minimum requirements or recommendations for criminal risk assessment reports across institutions, from which it can be concluded that the published quality standards influenced the judicial ordering practice in the direction of an increasing differentiation of the questions (see Chapter 3). However, exclusively considering the judicial questions posed to the experts after 2006, the differentiation of the formulated orders, even after the publication of the minimum requirements, is still heterogeneous. Overall, there was a stronger orientation of the judicial orders towards the questions specified in the minimum requirements, but the subsequent practice of judicial ordering continued to be heterogeneous and did not show any consistent prognostic questions. This was verified again when analyzing the exact wording of the judicial orders. Thus, hardly any consistent formulations of prognostic questions could be determined in the present sample. The results seem to be important, especially because significant correlations were found between the consideration of the prognostic questions in judicial orders and the answers given by the experts in assessment reports. The results of the present study show that the judicial order for a risk assessment report has an influence on the

answering of the fundamental prognostic questions and can thus significantly increase the quality of risk assessments in the sense of a better usability in the proceedings. Thus, the results show potential for quality assurance not only for the experts but also for the orderers of assessment reports. The methodological minimum requirements for criminal risk and responsibility assessment reports should not only be implemented by experts but should be considered by orderers – amongst others also in the context of a transparent presentation of the prerequisites for mandatory treatment in verdicts containing criminal responsibility assessments. The examination of the information on procedural outcomes in this thesis revealed the heterogeneity of judicial consideration of expert findings (see Chapter 4). In the majority of cases, the expert's findings were merely reproduced and evaluated in only ten percent of the cases, showing no discussion of the expert findings in most cases. In the current study, there were no significant differences over time, which seems to indicate that the methodological minimum requirements do not have a decisive influence on this judicial practice. In terms of general considerations of quality assurance in the field of criminal responsibility assessment, the provision of transcripts of judicial verdicts according to article 475 (4) of the German penal code could be helpful for experts to improve their ability to communicate psycho-scientific findings to the court regarding knowledge of the consideration of the assessment report in verdicts (Pfister, 2019).

Regarding the **degree of standardization in criminal assessments reports in professional practice**, the study results show a generally frequent usage of a diverse range of psychological tests in forensic-clinical practice, confirming the general acceptance within the forensic field (see Chapter 5). In risk assessment reports, psychological tests (mostly addressing personality traits) were used significantly more often (and in a significantly higher number) than in criminal responsibility assessment reports (primarily using intelligence and cognitive measures). While the usage of psychological testing increased significantly in risk assessment reports over time consistent with previous studies (Neal & Grisso, 2014; Serafim

et al., 2015; Wright et al., 2017), no significant difference could be confirmed regarding criminal responsibility reports. This is all the more noteworthy, as criminal responsibility assessments already had a lower level of test usage to begin with and it is legally mandatory to diagnose a psychological condition in a first step. However, in this German sample, psychological testing seemed to be more common and furthermore increasing in risk assessment reports than in criminal responsibility reports in recent years, documenting a growing relevance of interdisciplinarity in risk assessment in general. In more than a half of the final conclusions of risk assessment reports, psychological test results were not just reported generally but even considered and discussed in terms of risk and protective factors regarding criminal risk of individuals charged or convicted of offenses, underlining the importance of psychological testing in criminal risk assessments in real practice. All in all, the present results provide empirical data for the wide acceptance as well as variety and heterogeneity of test usage in forensic evaluations, as there seems to be no consensus about psychological test usage and the degree of formalisation in forensic evaluations (as reported in Archer et al., 2006; Golden & Lashley, 2014; Gowensmith & McCallum, 2019; Lally, 2003; Richards et al., 2015).

Regarding the degree of standardisation and structuring of criminal risk assessments, the results show that actuarial or clinical-structured risk assessment instruments were only used in less than half of the total sample or assessment reports (see Chapter 6). Despite the impressive evidence in the forensic research literature of the higher predictive accuracy of standardized risk assessment compared to clinical-unstructured, intuitive, or predominantly or exclusively experience-based approaches (see, e.g., Rettenberger, 2018, 2019; Viljoen et al., 2021), a comparatively high proportion of so-called "clinical" (i.e., unstructured, intuitive, and experience-based) predictions was still evident in the present sample (see Chapter 6 and 7). The results were in line with previous empirical analyses of risk assessment reports (e.g., Haubner-Mclean & Eher, 2014) and supported previous survey results on criminal risk

assessment practice in Germany (Rettenberger at al., 2017), which also indicated a use of standardized risk assessment instruments in approximately half of the risk assessments. According to surveys (Etzler & Rettenberger, 2019; Gregório Hertz et al., 2019), the standardization of criminal risk assessment seems to be more advanced in forensic institutions (social therapy, aftercare) than in external assessment practice. This could also be confirmed in the present study, as there seems to be still a relatively heterogeneous application practice of risk assessment methods in external assessment reports. Furthermore, the results of the present study show that the use of actuarial and clinical-structured risk assessment instruments appears to be partly dependent on examinee-related characteristics. Thus, it can be stated that in German-speaking risk assessment practice, several internationally established risk assessment instruments whose predictive validity is considered to be empirically validated, are applied regularly. However, there still seems to be a need for optimization, especially with regard to the use of actuarial instruments, which are empirically particularly well-validated (e.g. Hanson & Morton-Bourgon, 2009; Rettenberger, Rice, Harris & Eher, 2017).

Concerning the formalisation of risk communication in risk assessment reports, present results revealed a clear preference for a categorical form of communication (see Chapter 6), which is in line with international studies (Heilbrun et al., 2016; Singh et al., 2014) and those of the German-speaking area (Rettenberger at al., 2017). In just under two-thirds of the total sample, risk was communicated categorically or dichotomously. However, in just under one-third - particularly in the sub-sample of the university department - a combined form of communication could be identified, as it has been increasingly proposed or called for by authors in the recent past (including de Vogel et al., 2020; Eher et al., 2019; Hanson et al., 2017a; 2017b). The predominantly categorical risk communication could also be attributed to the more frequent use of clinical-structured risk assessment instruments

compared to actuarial tools in the present sample, as the nominal form of communication is the method of choice for the former group (von Franqué, 2013).

Regarding the **superiority of structured risk assessment approaches**, the results provide an empirical comparison of the predictive accuracy of unstructured and structured methodological approaches of risk assessment in a sample of German risk assessment reports (see Chapter 7). On the basis of the ongoing controversial debate on the relevance and benefits of intuitively and subjectively made assessment judgments (e.g., Dressel & Farid, 2018; Lin et al., 2020; Melton et al., 2018; see Chapter 1.2.1.2), more than half of the reports examined in the present study were considered as UCJs. In accordance with previously published results and meta-analytic findings (e.g., Ægisdóttir et al., 2006; Bengtson & Långström, 2007; Turgut et al., 2006), the comparisons presented here indicate a higher predictive accuracy for structured compared to unstructured risk assessment approaches for the prediction of general, violent, and sexual recidivism across different subgroups. When using the total follow-up time (average follow-up of 7.08 years) as well as when using the fixed 5-year follow-up time periods, both methods for calculating effect sizes (AUC and ORs/RRs) yielded a comparable result pattern in terms of a limited predictive accuracy of UCJs. Thereby, no clearly distinct superiority for either actuarial or SPJ approaches was identified; this finding confirmed similar results reported in previous studies which compared both methods (e.g., French, 2021; Hanson & Morton-Bourgon, 2009; Nicholls et al., 2013). A further noteworthy result refers to the finding that the combination of ARAIs and SPJ instruments yielded particularly high effect sizes for the prediction of (general, violent, and sexual) recidivism, which were predominantly higher than the predictive accuracy of ARAIs or SPJ instruments alone. This finding supports the idea of a convergent risk assessment approach where the strengths of both ARAIs and SPJ instruments are combined (Singer et al., 2016).

Before concluding with practical suggestions for those assessing criminal risk and/or responsibility, further unaddressed issues, limitations of the present thesis, and the need for further research will first be outlined.

8.2 Limitations and Need for Further Research

The present thesis naturally shows methodological limitations. Above all, the present thesis examined the minimum requirements and thus only one of many possible measures for quality assurance of assessment reports. A final assessment of quality was not the aim of this thesis and would exceed the limits of an operationalized set of criteria for methodological minimum requirements. For these purposes, other measures of quality assurance – in addition to standardization of criminal assessment practice - must be pursued. For example, a similar quality-assurance procedure to the peer review process applied to expert witness reports in family law (Kannegießer, 2018) would also be conceivable and desirable for criminal responsibility reports (Banse, 2017). Regarding this critical discussion, a further examination of the methodological minimum requirements also seems useful. In addition to new empirical findings, changes in the legal framework could also be considered and mapped. Similarly, the recommendations for risk assessment reports have recently been updated (Boetticher et al., 2019; Kroeber et al., 2019).

It is also important to note that we defined the degree of structure of the assessment reports by considering if and what underlying type and generation of risk assessment instrument (ARAIs vs. SPJ instruments) was used, i.e., if an ARAI or SPJ tool was used, assessments were considered as structured. Of course, this procedure could be interpreted as only a rough approximation to a sophisticated categorization of risk assessment approaches. Furthermore, there were standardised assessment procedures like the PCL-R, which can only hardly be allocated to the before mentioned groups of risk assessment methods.

It should also be noted that the methodological minimum requirements do not represent binding legal criteria, the non-implementation of which would constitute a legal

defect in any case. In this respect, it is not necessarily a quality defect in the context of the present study if experts do not follow the criteria. The recommendations did also not constitute binding legal criteria, so not meeting them did not display a failure of law. It was explicitly pointed out that experts could deviate from the criteria if there were objective reasons for doing so. Nevertheless, formulation of the minimum requirements for example already illustrates that they should be conducted regularly. Thus, no particularly high-threshold quality standards are described. The criteria reflect the minimum level of differentiation in the assessment reports, from which deviations should only be made in factually substantiated exceptional cases. Conversely, compliance with the methodological minimum requirements does not guarantee a "correct" assessment and expert (Konrad, 2010), especially since the minimum requirements are also criticized alongside concerns with operationalization (Eisenberg, 2005) and clinical and practical applicability for assessing the severity of diverse mental disorder groups (Konrad, Huchzermeier & Rasch, 2019; Dobbrunz & Briken, 2020; Fuß et al., 2020). The revised recommendations for risk assessment reports (Boetticher et al., 2019, Kröber et al. 2019) also listed that there was no claim to binding force, but that the expert would have to explain if he or she deviated from the principles for professional reasons in individual cases. In this respect, it also cannot necessarily be said that there is a lack of quality if experts do not use standardized instruments. Generally applicable guidelines or recommendations for the use of special risk instruments cannot be formulated and always depend on the context and the specific prognostic question. In principle, the selection of the appropriate methodological approach for the risk assessment is the responsibility of the expert. Nevertheless, it follows already from the formulation of the methodological minimum requirements or recommendations that statistical methods should be used regularly (Kröber et al., 2019). Empirical research results also impressively underline the added value of standardized assessment methodology.

Moreover, some criteria of the methodological minimum requirements were difficult to assess retrospectively based on the written assessment reports. However, the methodological limitations of the non-validated survey instrument due to the lack of psychometric quality and the restrictive three-level scaling of the questionnaire are offset by numerous possibilities regarding a precise differentiation, that allow operationalization of the quality of the risk assessment reports as applied in this study. The reliability criterion, on the other hand, was met by calculating the interrater reliability, as done so for all retrospective ratings. Reference has also been made to the limited comparability of the content of the two institutions emphazising the special status of institutions such as universities. In interpreting these findings, it is important to note that most assessment reports were conducted at a university hospital. Therefore, the standards of implementing methodological minimum requirements or risk assessment instruments and capacities of inclusion of psychological testing described may be different to those of clinicians who are not members of scientific forensic organizations like university hospitals or other research institutions. This was also reflected not only in the use of risk assessment instruments, but also in the involvement of additional psychological test examinations and social science explorations as part of a generally more pronounced interdisciplinary approach at the university institution. While experts at the university department processed risk assessment reports on individuals of both forensic psychiatric hospitals and the penitentiary, the sub-sample of the penitentiary only contained risk assessment reports on prisoners and persons in preventive detention regarding article 66 of the German penal code. Consequently, not only did the judicial orders and the imprisonment or accommodation of the examinees differ, so did the prevalence of (psychiatric) diagnoses according to ICD-10. Furthermore, assessment reports were contributed by a merely small number of experts, so individual preferences and opinions of experts e.g., on validity of assessment instruments or psychometric tests had an influence on present results. Moreover, it is not possible to ascertain whether assessors implementing

methodological minimum requirements or using validated, empirically supported risk assessment instruments are maybe in general more experienced or trained in conducting forensic evaluations compared to experts who do not use these methods. Confounding variables of assessors themselves may further explain differences in accuracy of risk assessments, so it may be a promising approach to control for different assessor characteristics in futures studies. Nonetheless, the integrated institutions represent common general forensic-clinical practice in Germany and the results were cross-institutional, from which generalizable conclusions can be drawn. It could therefore be concluded that a tendency towards standardized judgment based on risk assessment instruments in criminal risk assessments and implementation of methodological minimum requirements are increasing not only in university assessment practice but in general. However, the results indicate the need for further discussion about standardisation in clinical practice in terms of inclusion and heterogeneity of psychometric tests, risk assessment instruments, and methodological approaches in the field of forensic psychiatry and psychology, including more research on non-scientific institutions.

Furthermore, studies on the validity of risk assessment reports are likewise subject to limited interpretation, since the legal consequences of a negatively directed prognosis often lead to lengthy custodial measures, making it impossible to prove false-positive assessments (Dittmann, 2012). Another data-related limitation was that the information about recidivism was only based on one single data source (i.e., officially registered reconvictions in accordance to the Federal Central Register), which is inevitably an underestimation of the actual degree of repeated delinquency. Federal Central Register excerpts represent only one of several possible recidivism data sources.

Methodologically, experimental designs including a random assignment to a specific approach and the use of prospective-longitudinal field study designs would be the best way for (cross-)validating risk assessment approaches and instruments (e.g., Nitsche et al., 2022;

Rettenberger et al., 2017). However, particularly experimental research approaches can be accompanied by serious ethical conflicts and legal issues. Given the fact that one main aim of the present thesis was to identify clinically applied risk assessment methods in their everyday practice, the retrospective design was indispensable. But even if the present findings basically underline the importance and the relatively high predictive validity of structured risk assessment approaches, the need for further research became also clear, particularly about direct comparisons of the predictive accuracy of different methodological approaches in applied risk assessment settings.

8.3 Conclusion and Implication for Practice

Summarising, the results showed an increasing implementation of minimum requirements and standardization of clinical and judicial criminal risk assessment practice in terms of use of psychometric tests and risk assessment instruments. The results seem to be especially noteworthy, as it could be shown that compliance with the methodological minimum requirements and the degree of standardization or structuring of criminal risk assessment are associated with a higher accuracy of the risk assessments. Furthermore, the results indicated the importance of forensic psychological testing for prognostic judgements. On the one hand, the results indicate the presented (partial) positive effects, on the other hand, more efforts are needed regarding further quality assurance of criminal risk and responsibility assessments, as quality and formalisation in general practice still seem to be very heterogenous. Therefore, the present thesis identified potential for improvement not just for experts but also for orderers in terms of standardized practice in the assessment of both criminal risk and responsibility, as the judicial order for a risk assessment report had an influence on the quality of given prognostic answers in the reports and therefore on the usability in the proceedings.

On the basis of the ongoing controversial debate on the relevance and benefits of intuitively and subjectively made assessment judgments, the comparisons presented in the

thesis indicate a higher predictive accuracy for structured compared to unstructured risk assessment approaches for the prediction of general, violent, and sexual recidivism across different subgroups and follow-up periods. Given the weight of evidence in support of structured risk assessment, the findings underline the limited accuracy of UCJs and provide further support for the use of structured and standardized risk assessment procedures in the field of forensic sciences and psychological assessment of crime, delinquency, and recidivism. However, a comprehensive and convergent risk assessment approach including clinical-idiographic evaluation of the individual biography and background is not only mandatory with regard to the German penal law but yielded also the highest effect sizes for predictive accuracy (compared to ARAIs and SPJ instruments alone). As different previous studies and reviews indicated that the predictive accuracy of unstructured assessments could be possibly better as originally stated given the variety of methodological limitations of the existing research data (e.g., Lin et al., 2020; Melton et al., 2018; Viljoen et al., 2021), the present study remedied this research gap by overcoming at least some of the discussed methodological limitations of prior research.

Taken together, from an applied risk assessment perspective, the findings of the present study indicate, first, that both assessors and orderers should be encouraged more consistently to consider and apply the methodological minimum requirements, as those are associated with a higher accuracy of risk assessments and judicial orders do have an influence on the quality of assessment reports. Second, assessors should use indeed both actuarial and SPJ risk assessment approaches and combine and integrate the results of both into the final judgment. Irrespective of the fact that German case law stipulates a high degree of individualization of risk assessment (Boetticher et al., 2019), standardized risk assessment instruments provide empirically sound statements that (should) be an essential decision support in the required individual case assessment. Third, the competition of ARAIs and SPJ representatives in the past seems to be resolved in the present tense: The crucial issue is not

which approach is better but rather how to combine both in order to achieve the best prediction. The combination of both approaches is also relevant from another point of view: The use of ARAIs alone has inevitable limitations because the actuarial approach is not able to provide an individual (idiographic) explanatory model of delinquency. This means that a clinical-idiographic approach is necessary and even mandatory in the German penal law which is also confirmed by currently published recommendations for criminal risk assessment (Kröber et al., 2019). It is important to emphasize that a comprehensive and individually based risk assessment approach is therefore indispensable and the pure use of group-based statistics or nomothetic rules is insufficient (Ægisdóttir et al., 2006). However, individual clinical and case-specific risk assessments of examinees should though not be conducted intuitively but should always follow strictly scientific, transparent, and evidence-based standards. Therefore, current comprehensive risk assessment models have tried to integrate the individual and relatively flexible approach of case formulation with the standardized assessment approach of ARAIs and SPJ instruments (e.g., the Case Formulation Incorporating Risk Assessment [CAFIRA] model; Craig & Rettenberger, 2018; Craig et al., 2020). Consequently, presented scientific findings in risk assessment research should also be integrated in criminal responsibility assessment practice.

References

Adams, B. (2009). *Die Qualität forensischer Gutachten über die Legalprognose von Straftätern* [The quality of criminal risk assessment reports] [Master´s Thesis, Albert-Ludwigs-University of Freiburg i. Br.]. Forschungsdatenbank Albert-Ludwigs-Universität Freiburg. https://www.psychologie.uni-freiburg.de/forschung/Forschungsdatenbank/sozpsy/arbeiten

Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The meta-analysis of Clinical Judgment Project: Fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist, 34*(3), 341–382. https://doi.org/10.1177/0011000005285875

Alex, M. (2010). *Nachträgliche Sicherungsverwahrung – ein rechtsstaatliches und kriminalpolitisches Debakel* [Subsequent preventive detention - a constitutional and criminal policy debacle]. Felix-Verlag.

Allison, P. D. (1984). *Event history analysis: Regression for longitudinal event data.* Sage Publications. http://doi.org/10.4135/9781412984195

American Psychological Association (2013). Specialty guidelines for forensic psychology. *The American Psychologist, 68*(1), 7–19. https://doi.org/10.1037/a0029889

Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised.* Multi-Health Systems.

Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency, 52*(1), 7–27. https://doi.org/10.1177%2F0011128705281756

Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07

Banse, R. (2017). Qualitätssicherung von rechtspsychologischen Sachverständigengutachten durch ein moderiertes Peer-Reviewverfahen: Ein Vorschlag zur Diskussion [Quality assurance of expert legal psychological assessment reports through a moderated peer review process: a proposal for discussion]. *Praxis der Rechtspsychologie, 27*(2), 113–130.

Basdekis-Jozsa, R., Rettenberger, M., & Briken, P. (2013). Begutachtung bei sexuellen Störungen [Assessment of individuals with sexual disorders]. In P. Briken & M. Berner (Eds.), *Praxisbuch Sexuelle Störungen* (pp. 257–271). Thieme.

Belschner, W., Lischke, G., & Selg, H. (1971). *Foto-Hand-Test (FHT): Zur Erfassung der Aggressivität* [Photo-Hand-Test: For the assessment of aggressiveness]. Karl Alber.

Bengtson, S., & Langström, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: A Journal of Research and Treatment, 19*(2), 135–153. https://doi.org/10.1007/s11194-007-9044-5

Blair, P. R., Marcus, D. K., & Boccaccini, M. T. (2008). Is there an allegiance effect for assessment instruments? Actuarial risk assessment as an exemplar. *Clinical Psychology: Science and Practice, 15*(4), 346–360. https://doi.org/10.1111/j.1468-2850.2008.00147.x

Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic test usage by forensic psychologists in emotional injury cases. *Professional Psychology: Research and Practice, 30*(3), 253–259. https://doi.org/10.1037/0735-7028.30.3.253

Boer, D. P., Hart, S. D., Kropp, P. R., & Webster C. D. (1997). *Manual for the Sexual Violence Risk – 20. Professional Guidelines for Assessing Risk of Sexual Violence*. The Mental Health, Law, and Policy Institute, Simon Fraser University.

Bonta, J., Law, M., & Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin,*

123(2), 123–142. https://doi.org/10 .1037/0033-2909.123.2.123

Böhm, K. (2018). Zuständigkeiten und Aufgaben der Kriminalprognostik unter besonderer Berücksichtigung der Mindestanforderungen für Prognosegutachten [Responsibilities and tasks of criminal risk assessment with special regard to methodological minimum requirements for risk assessment reports]. *Recht & Psychiatrie, 36*(3), 133–138.

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 [Recommendations for criminal risk assessment reports: Legal framework for risk assessment reports in criminal proceedings]. *Neue Zeitschrift für Strafrecht, 39*(10), 553–573. https://doi.org/10.1007/s11757-019-00557-0

Boetticher, A., Kröber, H. L., Müller-Isberner, R., Böhm, K. M., Müller-Metz, R., & Wolf, T. (2006). Mindestanforderungen für Prognosegutachten [Methodological minimum requirements for criminal risk assessment reports]. *Neue Zeitschrift für Strafrecht, 26*(10), 537–544. https://doi.org/10.1007/s11757-007-0017-9

Boetticher, A., Kröber, H. L., Müller-Isberner, R., Böhm, K. M., Müller-Metz, R., & Wolf, T. (2007). Mindestanforderungen für Prognosegutachten [Methodological minimum requirements for criminal risk assessment reports]. *Forensische Psychiatrie, Psychologie, Kriminologie, 1*(2), 90–100. https://doi.org/10.1007/s11757-007-0017-9

Borum, R., & Grisso, T. (1995). Psychological test use in criminal forensic evaluations. *Professional Psychology: Research and Practice, 26*(5), 465–473. https://doi.org/10.1037/0735-7028.26.5.465

Campbell, M. A., French, S., & Gendreau, P. (2009). The prediction of violence in adult offenders: A meta-analytic comparison of instruments and methods of assessment. *Criminal Justice and Behavior, 36*(6), 567–590. https://doi.org/10.1177/0093854809333610

Costa, P. T., & McCrae, R. R. (2008). The revised NEO personality inventory (NEO-PI-R). In G. J. Boyle & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 2 - personality measurement and testing* (pp. 179–198). SAGE Publications Inc. https://doi.org/10.4135/9781849200479.n9

Craig, L. A., & Rettenberger, M. (2018). An etiological approach to sexual offender assessment: CAse Formulation Incorporating Risk Assessment (CAFIRA). *Current Psychiatry Reports, 20*(6), 43. https://doi.org/10.1007/s11920-018-0904-0

Craig, L. A., Rettenberger, M., & Beech, A. R. (2020). Neurobiological considerations on the etiological approach to sexual offender assessment: CAse Formulation Incorporating Risk Assessment–version 2 (CAFIRA–v2). In J. Proulx, F. Cortoni, L. A. Craig & E. J. Letourneau (Eds.), *The Wiley handbook of what works with sexual offenders: Contemporary perspectives in theory, assessment, treatment, and prevention* (pp. 153–173). Wiley-Blackwell.

Cutler, B. L., & Kovera, M. B. (2011). Expert psychological testimony. *Current Directions in Psychological Science, 20*(1), 53–57. https://doi.org/10.1177/0963721410388802

Dahle, K. P. (2005a). *Psychologische Kriminalprognose. Wege zu einer integrativen Methodik für die Beurteilung der Rückfallwahrscheinlichkeit bei Strafgefangenen* [Psychological criminal risk assessment. Towards an integrative methodology for assessing the probability of recidivism in prisoners.]. Centaurus.

Dahle, K. P. (2005b). Psychologische Begutachtung zur Kriminalprognose [Psychological assessment for criminal risk assessment]. In H. L. Kröber & M. Steller (Eds.), *Psychologische Begutachtung im Strafverfahren* (pp. 215–219). Steinkopff.

Dahle, K. P. (2006). Grundlagen und Methoden der Kriminalprognose [Basic principles and methods of criminal risk assessment] In H. L. Kröber, D. Dölling, N. Leygraf & H. Sass (Eds.), *Handbuch der Forensischen Psychiatrie: Psychiatrische Kriminalprognose und Kriminaltherapie* (pp. 1–67). Steinkopff.

Dahle, K. P. (2010). Die Begutachtung der Gefährlichkeits- und Kriminalprognose des Rechtsbrechers [The criminal risk assessment of offenders]. In R. Vollbert & K. P. Dahle (Eds.), *Forensisch-psychologische Diagnostik im Strafverfahren* (pp. 67–114). Hogrefe.

Dahle, K. P., Bliesener, T., Gretenkord, L., & Schwabe-Höllein, M. (2012). Qualitätssicherung in der forensischen Psychologie [Quality assurance in forensic psychology – The certified advanced training as a specialist psychologist for legal psychology of the BDP/DGPs]. *Forensische Psychiatrie, Psychologie, Kriminologie, 6*(4) 243–249. https://doi.org/10.1007/s11757-012-0181-4

Dahle, K. P., Harwardt, F., & Schneider-Njepel, V. (2012). *Inventar zur Einschätzung des Rückfallrisikos und des Betreuungs-und Behandlungsbedarfs von Straftätern. Deutsche Version des Level of Service Inventory-revised nach Don Andrews and James Bonta* [Inventory for the assessment of the recidivism risk and the care and treatment needs of offenders. German version of the Level of Service Inventory-revised from Don Andrews and James Bonta]. Hogrefe.

Dahle, K. P., & Lehmann, R. J. B. (2018). Zum prognostischen Mehrwert einer integrativen nomothetisch-idiografischen kriminalpsychologischen Prognosebeurteilung – Eine empirische Untersuchung an männlichen Gewalt- und Sexualstraftätern [On the prognostic added value of an integrative, nomothetic-idiographic criminal psychological risk assessment – An empirical study on male violent and sexual offenders]. *Forensische Psychiatrie, Psychologie, Kriminologie, 12*(2), 37–50. https://doi.org/10.1007/s11757-017-0462-z

Dahle, K. P., Schneider-Njepel, V., & Ziethen, F. (2009). Methodische Möglichkeiten und Grenzen psychologischer Kriminalprognosen. [Methodological possibilities and limitations of psychological criminal risk assessments]. In S. Dauer, R. Doberenz, C. Orth & G. Teichert (Eds.), *Rechtspsychologie zwischen Politik, Justiz und Medien*

[Legal psychology between politics, judiciary, and the media] (pp. 69–93). Pabst.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845. https://doi.org/10.2307/2531595

de Vogel, V., Bosker, J., & van den Broek, E. (2020). Forensische Risikokommunikation [Forensic risk communication]. *Forensische Psychiatrie und Psychotherapie, 27*(3), 326–342.

de Vogel, V., de Ruiter, C., Hildebrand, M., Bos, B., & van de Ven, P. (2004). Type of discharge and risk of recidivism measured by the HCR-20: A retrospective study in a Dutch sample of treated forensic psychiatric patients. *The International Journal of Forensic Mental Health, 3*(2), 149–165. https://doi.org/10.1080/14999013.2004.10471204

de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive Validity of the SVR-20 and Static-99 in a Dutch Sample of Treated Sex Offenders. *Law and Human Behavior, 28*(3), 235–251. https://doi.org/10.1023/B:LAHU.0000029137.41974.eb

de Vogel, V., de Vries Robbé, M., de Ruiter, C., & Bouman, Y. H. A. (2011). Assessing protective factors in forensic psychiatric practice: Introducing the SAPROF. *International Journal of Forensic Mental Health, 10*(3), 171–177. https://doi.org/10.1080/14999013.2011.600230

Diagnostik- und Testkuratorium. (2018). TBS-DTK. Testbeurteilungssystem des Diagnostik- und Testkuratoriums der Föderation Deutscher Psychologenvereinigungen. Revidierte Fassung vom 03. Jan. 2018 [Test assessment system of the diagnostic and test curatorship of the federation of German psychological associations. Revised version January 3rd., 2018]. *Psychologische Rundschau, 69*, 109–116. https://doi.org/10.1026/0033-3042/a000401

Dittmann, V. (2000). Was kann die Kriminalprognose heute leisten? [What can criminal risk

assessment achieve today?] In S. Bauhofer, P. H. Bolle & V. Dittmann (Eds.), *Gemeingefährliche Straftäter* (pp. 66–95). Rüegger.

Dittmann, V. (2012). Lieber würfeln oder eine Münze werfen? Über Risiken, wissenschaftliche Erkenntnisse und Prognosen. [Better to roll the dice or flip a coin? On risks, scientific findings and risk assessments]. In E. Yundina, S. Stübner, M. Hollweg & C. Stadtland (Eds.), *Forensische Psychiatrie als interdisziplinäre Wissenschaft: Festschrift zum Geburtstag von Norbert Nedopil* (pp. 45–58). Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Dobrunz, S., & Briken, P. (2020). Operationalisierte, kriteriengeleitete Beurteilung der Schuldfähigkeit bei paraphilen Störungen – Stand und Ausblick [Operationalized, criterion-referenced assessment of the criminal responsibility in paraphilic disorders – status and outlook]. *Forensische Psychiatrie, Psychologie, Kriminologie, 14*, 430–436. https://doi.org/10.1007/s11757-020-00633-w

Douglas, K. S., Hart, S. D., Webster, C. D., Belfrage, H., Guy, L. S., & Wilson, C. M. (2014). Historical-Clinical-Risk Management-20, Version 3 (HCR-20V3): Development and overview. *The International Journal of Forensic Mental Health, 13*(2), 93–108. https://doi.org/10.1080/14999013.2014.906519

Douglas, K. S., Yeomans, M., & Boer, D. P. (2005). Comparative validity analysis of multiple measures of violence risk in a sample of criminal offenders. *Criminal Justice and Behavior, 32*(5), 479–510. https://doi.org/10.1177/0093854805278411

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances, 4*(1), 1–5. https://doi.org/10 .1126/sciadv.aao5580

Dunn, E., Felthous, A. R., Gagné, P., Harding, T., Kaliski, S., Kramp, P., Lindqvist, P., Nedopil, N., Ogloff, J. R. P., Skipworth, J., Taylor, P. J., Thomson, L., & Yoshikawa, K. (2014). *Forensic psychiatry and its interfaces outside the UK and Ireland*. In J. Gunn & P. J. Taylor (Eds.), *Forensic Psychiatry. Clinical, Legal and Ethical Issues*

(2nd ed.). (pp. 111–147). Taylor & Francis Group. https://doi.org/10.1201/b15462

Eher, R., Rettenberger, M., Etzler, S., Eberhaut, S., & Mokros, A. (2019). Eine gemeinsame Sprache für die Risikokommunikation bei Sexualstraftätern – Trenn- und Normwerte für das neue Fünf-Kategorienmodell des Static-99 [A common language for communicating risk in sexual offenders – the 5-category model Static-99]. *Recht & Psychiatrie, 37*(2), 91–99.

Eisenberg, U. (2005) *Kriminologie* [Criminology] (6th ed.). Beck.

Engel, R. R., Hathaway, S. R., & McKinley, J. C. (2000). *Minnesota Multiphasic Personality Inventory–2*. Huber.

Etzler, S., Eher, R., & Rettenberger, M. (2020). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the Stable-2007. *Assessment, 27*(4), 822–839. https://doi.org/10.1177/1073191118754705

Etzler, S., & Rettenberger, M. (2019). Psychologische Diagnostik im Rahmen der Behandlung von Gewalt- und Sexualstraftätern im Justizvollzug: Eine Vollerhebung diagnostischer Praxis der sozialtherapeutischen Einrichtungen in Deutschland [Psychological assessment for incarcerated violent and sexual offenders: A complete survey of the diagnostical practice in social therapy units in Germany]. *Diagnostica, 66*(1), 14–24. https://doi.org/10.1026/0012-1924/a000235

Fegert, J. M., Häßler, F., Schnoor, K., Reberning, E., König, C., Auer, U., & Schläfke, D. (2003). *Bestandsaufnahme und Qualitätssicherung der forensisch-psychiatrischen Gutachtertätigkeit in Mecklenburg-Vorpommern bei Tötungs- und Brandstiftungsdelikten. Abschlussbericht des Landesmodellprojektes* [Inventory and quality assurance of the forensic psychiatric expert witness report practice in Mecklenburg-Vorpommern in cases of homicide and arson offenses. Final report of the state model project]. Books on Demand.

Ferguson, G. E., Eidelson, R. J., & Witt, P. H. (1998). New Jersey's Sex Offender Risk

Assessment Scale: Preliminary Validity Data. *The Journal of Psychiatry & Law, 26*(3), 327–349. *https://doi.org/10.1177/009318539802600303*

French, S. A. (2021). *What works! A Meta-Analysis of general risk prediction instruments for adult criminal offenders* [Dissertation, University of New Brunswick Saint John]. University of New Brunswick UNB Scholar. https://unbscholar.lib.unb.ca/islandora/object/unbscholar%3A10485

Fuger, K. D., Acklin, M. W., Nguyen, A. H., Ignacio, L. A., & Gowensmith, W. N. (2014). Quality of criminal responsibility reports submitted to the Hawaii judiciary. *International Journal of Law and Psychiatry, 37*(3), 272–280. https://doi.org/10.1016/j.ijlp.2013.11.020

Fuß, J., Voulgaris, A., & Briken, P. (2020). Schuldfähigkeit bei Personen, die wegen Nutzung von Missbrauchsabbildungen angeklagt sind [Criminal responsibility in individuals convicted of the use of abuse images]. *Forensische Psychiatrie, Psychologie, Kriminologie, 14*, 437–445. https://doi.org/10.1007/s11757-020-00624-x

Glancy, G. D., Ash, P., Bath, E. P., Buchanan, A., Fedoroff, P., Frierson, R. L., Harris, V. L., Hatters Friedman, S. J., Hauser, M. J., Knoll, J., Norko, M., Pinals, D., Price, M., Recupero, P., Scott, C. L., & Zonana, H. V. (2015). AAPL Practice Guideline for the Forensic Assessment. *The Journal of the American Academy of Psychiatry and the Law*, *43*(2, Suppl.), S3–S53. https://pubmed.ncbi.nlm.nih.gov/26054704/

Golden, C. J., & Lashley, L. (2014). *Forensic neuropsychological evaluation of the violent offender*. Springer International Publishing. https://doi.org/10.1007/978-3-319-04792-8

Gowensmith, W. N., & McCallum, K. E. (2019). Mirror, mirror on the wall, who's the least biased of them all? Dangers and potential solutions regarding bias in forensic psychological evaluations. *South African Journal of Psychology*, *49*(2), 165–176. https://doi.org/10.1177/0081246319835117

Gregório Hertz, P., Breiling, L., Turner, D., & Rettenberger, M. (2019). Die Praxis der ambulanten Nachsorge für haftentlassene Sexualstraftäter in Deutschland [Outpatient aftercare of sexual offenders after release from prision in Germany]. *Recht & Psychiatrie, 37*(3), 157–164. https://krimpub.krimz.de/frontdoor/deliver/index/docId/333/file/Gregorio_Hertz_RuP_2019_3.pdf

Gregório Hertz, P., Eher, R., Etzler, S., & Rettenberger, M. (2021). Cross-validation of the revised version of the Violence Risk Appraisal Guide (VRAG-R) in a sample of individuals convicted of sexual offenses. *Sexual Abuse, 33*(1), 63–87. https://doi.org/10.1177/1079063219841901

Gregório Hertz, P., Müller, M., Barra, S., Turner, D., Rettenberger, M., & Retz, W. (2021). The predictive and incremental validity of ADHD beyond the VRAG-R in a high-risk sample of young offenders. *European Archives of Psychiatry and Clinical Neuroscience.* Advance online publication. https://doi.org/10.1007/s00406-021-01352-x

Gretenkord, L. (2013). Warum Prognoseinstrumente? [Why risk assessment instruments?] In M. Rettenberger & F. von Franqué (Eds.), *Handbuch kriminalprognostischer Verfahren* (pp. 19–36). Hogrefe.

Grisso, T. (2010). Guidance for improving forensic reports: A review of common errors. *Open Access Journal of Forensic Psychology*, *2*, 102–115. https://escholarship.umassmed.edu/psych_pp/282/

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://psycnet.apa.org/buy/1997-02834-005

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*(1), 19–30. https://doi.org/10.1037/1040-3590.12.1.19

Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A metaanalytic survey* [Doctoral dissertation, Simon Fraser University]. Simon Fraser University Summit. http://summit.sfu.ca/item/9247

Haarig, F., Blasé, T., & Sedlmeier, P. (2012). Vorhersage von Rückfälligkeit bei Sexualstraftätern. Wie gut sind die Gutachten und wie könnte man sie verbessern? [Predicting recidivism in sexual offenders. How good are the expert assessment reports and how could they be improved?] *Monatsschrift für Kriminologie und Strafrechtsreform, 95*(6), 392–412. https://doi.org/10.1515/mks-2012-950602

Habermeyer, E. (2005). Kriterienkataloge: Ein Beitrag zur Qualitätssicherung in der Forensischen Psychiatrie [Criteria catalogues: A contribution to quality assurance in the forensic psychiatry]. In F. Schneider (Ed.), *Entwicklungen in der Psychiatrie* (pp. 375–386). Springer.

Habermeyer, E. (2008). *Die Maßregel der Sicherungsverwahrung* [The preventive detention measure]. (Bd. 116). Steinkopff. https://doi.org/10.1007/978-3-7985-1844-5

Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., & Dougherty, E. R. (2010). Small-sample precision of ROC-related estimates. *Bioinformatics, 26*(6), 822–830. https://doi.org/10.1093/bioinformatics/btq037

Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology/Psychologie canadienne, 50*(3), 172–182. https://doi.org/10.1037/a0015726

Hanson, R. K., Babchishin, K. M., Helmus, L. M., Thornton, D., & Phenix, A. (2017a). Communicating the results of criterion referenced prediction measures: Risk categories for the Static-99R and Static-2002R sexual offender risk assessment tools. *Psychological*

Assessment, 29(5), 582–597. https://doi.org/10.1037/pas0000371

Hanson, R. K., Bourgon, G., McGrath, R. J., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017b). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language,* The Council of State Governments Justice Center.

Hanson, R. K., & Harris, A. (2002). *The Sex Offender Need Assessment Rating (SONAR): A method for measuring change in risk levels 2000-1*. Department of the Solicitor General of Canada. www.ps-sp.gc.ca/res/cor/rep

Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: a meta-analysis of 118 prediction studies. *Psychological Assessment, 21*(1), 1–21. https://doi.org/10.1037/a0014421

Hanson, R. K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior, 24*(1), 119–136. https://doi.org/10.1023/A:1005482921333

Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002 2003-01*. Public Works and Government Services Canada.

Hare, R. D. (2003). *Manual for the revised Psychopathy Checklist* (2nd ed.). Multi-Health Systems.

Harris, A., Phenix, A., Thornton, D., & Hanson, R. K. (2003). *STATIC-99: Coding Rules Revised – 2003.* Solicitor General.

Harris, G. T., Rice, M. E., & Quinsey, V. L. (2010). Allegiance or fidelity? A clarifying reply. *Clinical Psychology: Science and Practice, 17*(1), 82–89. https://doi.org/10.1111/j.1468-2850.2009.01197.x

Haubner-Maclean, T., & Eher, R. (2014). Nicht mehr gefährlich und doch rückfällig? Die ungenügende Abbildung gefährlichkeitsrelevanter Merkmale bei rückfälligen ehemals untergebrachten Sexualstraftätern [Not dangerous anymore but still relapsing? Flawed

risk assessment in sexual offenders released from a high security forensic hospital who subsequently relapsed]. *Recht & Psychiatrie, 32*(2), 69–79.

Heilbrun, K., Grisso, T., & Goldstein, A. (2009). *Foundations of forensic mental health assessment (Best practices in forensic mental health assessment).* Oxford University Press. https://psycnet.apa.org/record/2008-14685-000

Heilbrun, K., Marczyk, G., DeMatteo, D., & Mack-Allen, J. (2007). A principles-based approach to forensic mental health assessment: Utility and update. In A. M. Goldstein (Ed.), *Forensic Psychology: Emerging Topics and Expanding Roles*. (pp. 45–72). John Wiley & Sons. https://psycnet.apa.org/record/2006-13463-002

Heilbrun, K., Newsham, R., & Pietruszka, V. (2016). Risk communication: An international update. In S. Fazel, J. P. Singh & S. Björkly (Eds.). *International perspectives on risk assessment* (pp. 26–39). Oxford University Press.

Heilbrun, K., Yasuhara, K., & Shah, S. (2010). Violence risk assessment tools: Overview and critical analysis. In R. K. Otto & K. S. Douglas (Eds.), *Handbook of violence risk assessment* (pp. 1–17). Routledge/Taylor & Francis.

Helmus, L. M., & Babchishin, K. M. (2017). Primer on risk assessment and the statistics used to evaluate its accuracy. *Criminal Justice and Behavior, 44*(1), 8–25. https://doi.org/10.1177/0093854816678898

Herschbach, B. M. (2009). *Der Einfluss der Gesetzesänderung vom 26.10.1998 („Gesetz zur Bekämpfung von Sexualdelikten und anderen gefährlichen Straftaten) – Ein Vergleich vor und nach dieser Gesetzesänderung erstellter Prognosegutachten* [The influence of the amendment of the law of 26.10.1998 ("Law to Combat Sexual Offenses and Other Dangerous Crimes") - A comparison of risk assessment reports conducted before and after the amendment of law] [Dissertation, Ludwig-Maximilians University, Munich]. Elektronische Hochschulzeitschriften LMU. https://doi.org/10.5282/edoc.9967

Hörmann, H., & Moog, W, (1957). *Rosenzweig Picture Frustration Test/ Form für*

Erwachsene – deutsche Fassung [Rosenzweig Picture Frustration Test / Adults – German Version]. Hogrefe.

Jackson, R. L., Rogers, R., & Shuman, D. W. (2004). The accuracy and adequacy of sexually violent predator evaluations: Contextualised risk assessment in clinical practice. *International Journal of Forensic Mental Health, 3*(2), 115–129. https://doi.org/10.1080/14999013.2004.10471201

Johansen, S. H. (2006). Accuracy of predictions of sexual offense recidivism: A comparison of actuarial and clinical methods. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 68*(3-B), 1929.

Kammeier, H. (2018). Die Maßregeln. Rechtliche Grundlagen [Mandatory treatments. Legal basics]. In F. Schmidt-Quernheim & T. Hax-Schoppenhorst (Eds.), *Praxisbuch Forensische Psychiatrie: Behandlung und ambulante Nachsorge im Maßregelvollzug* (pp. 33–35). Hogrefe.

Kannegießer, A. (2018). Pilotprojekt Peer-Review-Verfahren für familienpsychologische Gutachten [Pilot peer review project for family psychology expert assessment reports]. *Praxis der Rechtspsychologie, 28*(2), 181–185.

Kinzig, J. (1997a). Die Gutachtenpraxis bei der Anordnung von Sicherungsverwahrung [Expert report practice when ordering preventive detention]. *Recht & Psychiatrie, 15*, 9–20.

Kinzig, J. (1997b). Die Praxis der Sicherungsverwahrung [The practice of preventive detention]. *Zeitschrift für die gesamte Strafrechtswissenschaft, 109*(1), 122–164. https://doi.org/10.1515/zstw.1997.109.1.122

Konrad, N. (2010). Schlechtachten trotz Einhaltung der Mindestanforderungen an Prognosegutachten [Low quality assessment reports despite compliance with the minimum requirements for risk assessment reports]. *Recht & Psychiatrie, 28*(1), 30–32.

Konrad, N., Huchzermeier, C., & Rasch, W. (2019). *Forensische Psychiatrie und Psychotherapie: Rechtsgrundlagen, Begutachtung und Praxis* [Forensic psychiatry and psychotherapy: legal basics, assessment, and practice] (5th ed). Kohlhammer.

Kröber, H. L. (2006). Praxis der kriminalprognostischen Begutachtung: handwerkliche Mindeststandards und kasuistische Illustration. [Criminal risk assessment practice: Manual minimum standards and casuistic illustration]. In H. L. Kröber, D. Dölling, N. Leygraf & H. Sass (Eds.) *Handbuch der Forensischen Psychiatrie: Psychiatrische Kriminalprognose und Kriminaltherapie* (pp. 173–192). Steinkopff.

Kröber, H. L., Brettel. H., Rettenberger, M., & Stübner, S. (2019). Empfehlungen für Prognosegutachten. Erfahrungswissenschaftliche Empfehlungen für kriminalprognostische Gutachten [Recommendations for criminal risk assessment reports: Empirically based recommendations for criminal risk assessment reports]. *Forensische Psychiatrie, Psychologie, Kriminologie*, *13*(4), 334–342. https://doi.org/10.1007/s11757-019-00558-z

Kunzl, F. K., Eher, R., Lamott, F., Mörtl, K., Modica, C., Rosenow, T., & Pfäfflin, F. (2009). Qualitätsanalyse forensisch-psychiatrischer Zurechnungsfähigkeits- und Prognosegutachten über Sexualstraftäter [Quality analysis of forensic psychiatric criminal responsibility and risk assessment reports of sexual offenders]. *PPmP – Psychotherapie · Psychosomatik · Medizinische Psychologie*, *59*(02), A051. https://doi.org/10.1055/s-0029-1208192

Kunzl, F., & Pfaefflin, F. (2011). Qualitätsanalyse österreichischer Gutachten zur Zurechnungsfähigkeit und Gefährlichkeitsprognose [Quality analysis of Austrian expert reports on criminal responsibility and risk assessment.]. *Recht & Psychiatrie*, *29*(3), 152–159.

Kury, H., & Adams, B. (2010). Prognosegutachten im Strafvollzug. [Risk assessment reports in the penal system]. *Forum Strafvollzug*, *59*(2), 81–87.

Kury, H., Brandenstein, M., & Riegl, M. (2009). Zur Bedeutung von externen Kriminalprognosen für Vollzugsentscheidungen [On the importance of external risk assessments for penal system decisions]. In Arbeitsgemeinschaft Strafrecht des Deutschen Anwaltsvereins (Ed.), *Strafverteidigung im Rechtsstaat. 25 Jahre Arbeitsgemeinschaft Strafrecht des Deutschen Anwaltsvereins* (pp. 976–1002). Nomos. http://hdl.handle.net/11858/00-001M-0000-002E-4443-B

Kury, H., & Obergfell-Fuchs, J. (2012). *Rechtspsychologie. Forensische Grundlagen und Begutachtung. Ein Lehrbuch für Studium und Praxis* [Legal psychology. Forensic basics and expert assessment. A textbook for study and practice]. Kohlhammer.

Kury, H., & Wertz, M. (2017). Zur Qualität forensisch-psychiatrischer und -psychologischer kriminalprognostischer Gutachten. Ergebnisse einer empirischen Untersuchung. [On the quality of forensic psychiatric and psychological risk assessment reports. Findings of an empirical examination]. In H. Pollähne & C. Lange-Joest (Eds.), *Achtung: Begutachtung! Sachverständige in Justiz und Gesellschaft: Erwartungen und Verantwortung* (pp. 115–152). LIT Verlag.

Lally, S. J. (2003). What tests are acceptable for use in forensic evaluations? A survey of experts. *Professional Psychology: Research and Practice, 34*(5), 491–498. https://doi.org/10.1037/0735-7028.34.5.491

Lees-Haley, P., Smith, H., Williams, C., & Dunn, J. (1996). Forensic neuropsychological test usage: An empirical survey. *Archives of Clinical Neuropsychology, 11*(1), 45–51. https://doi.org/10.1016/0887-6177(95)00011-9

Leonhart, R. (2004). *Lehrbuch Statistik: Einstieg und Vertiefung* [Textbook of statistics: Introduction and consolidation]. Hogrefe.

Leygraf, N. (2015). Die Begutachtung der Gefährlichkeitsprognose [Criminal risk assessment]. In H. Foerster & H. Dreßing (Eds.), *Psychiatrische Begutachtung* (pp. 413–426). Elsevier. http://dx.doi.org/10.1016/B978-3-437-22901-5.50029-6

Lin, Z. J., Jung, J., Goel, S., & Skeem, J. (2020). The limits of human predictions of recidivism. *Science Advances, 6*(7), eaaz0652. https://doi.org/10.1126/sciadv.aaz0652

Litwack, T. R. (2001). Actuarial versus clinical assessments of dangerousness. *Psychology, Public Policy, and Law, 7*(2), 409–443. https://doi.org/10.1037/1076-8971.7.2.409

Lodewijks, H. P., Doreleijers, T. A., de Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, *31*(3), 263–271. https://doi.org/10.1016/j.ijlp.2008.04.009

Matthes, A., & Rettenberger, M. (2008). *Die deutsche Version des Stable-2007. [The German version of the Stable-2007*]. Institut für Gewaltforschung und Prävention (IGF).

McLaughlin, J. L., & Kan, L. Y. (2014). Test usage in four common types of forensic mental health assessment. *Professional Psychology: Research and Practice, 45*(2), 128–135. http://dx.doi.org/10.1037/a0036318

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* University of Minnesota Press. https://doi.org/10.1037/11281-000

Melton, G. B., Petrila, J., Poythress, N. G., Slobogin, C., Otto, R. K., Mossman, D., & Condie, L. O. (2018). *Psychological evaluations for the courts: A handbook for mental health professionals and lawyers* (4th ed.). Guilford Press.

Meyer, G. J., & Kurtz, J. E. (2006). Advancing personality assessment terminology: Time to retire "objective" and "projective" as personality test descriptors. *Journal of Personality Assessment, 87*(3), 223–225. https://doi.org/10.1207/s15327752jpa8703_01

Meyer, G. J., Viglione, D. J., Mihura, J. L., Erard, R. E., & Erdberg, P. (2011). *Rorschach Performance Assessment System: Administration, coding, interpretation, and technical manual.* Rorschach Performance Assessment System LLC. https://r-pas.org/

Miller, L. (2013). Psychological evaluations in the criminal justice system: Basic principles

and best practices. *Aggression and Violent Behavior, 18*(1), 83–91. https://doi.org/10.1016/j.avb.2012.10.005

Mokros, A., Hollerbach, P., Nitschke, J., & Habermeyer, E. (2017). *Deutsche Version der Hare Psychopathy Checklist-Revised (PCL-R) [The German version of the Hare Psychopathy Checklist-Revised (PCL-R)]*. Hogrefe.

Mokros, A., Vohs, K., & Habermeyer, E. (2014). Psychopathy and violent reoffending in German-speaking countries: A meta-analysis. *European Journal of Psychological Assessment, 30*(4), 117–129. http://dx.doi.org/10.1027/1015-5759/a000178

Morey, L. C. (1991*). Personality assessment inventory: Professional manual*. Psychological Assessment Resources.

Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The thematic apperception test. *Archives of Neurology & Psychiatry, 34*(2), 289–306. https://doi.org/10.1001/archneurpsyc.1935.02250200049005

Mosbacher, A. (2020). Erfolgreiche Revisionen wegen vermeintlicher oder tatsächlicher Gutachtenfehler [Successful appeals of alleged or actual expert assessment report deficits]. *Forensische Psychiatrie, Psychologie, Kriminologie, 14*(4), 446–454. https://doi.org/10.1007/s11757-020-00629-6

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62*(4), 783–792. https://doi.org/10.1037/0022-006X.62.4.783

Müller, J. L., & Nedopil, N. (2017). *Forensische Psychiatrie: Klinik, Begutachtung und Behandlung zwischen Psychiatrie und Recht* [Forensic psychiatry: clinic, assessment and treatment between psychiatry and law]. Thieme.

Müller, J. L., & Stolpmann, G. (2015). Legalbewährung nach rechtskräftiger Ablehnung einer nachträglichen Anordnung der Unterbringung in der Sicherungsverwahrung [Legal probation after a legally binding rejection of a subsequent order of placement in

preventive detention]. *Monatsschrift für Kriminologie und Strafrechtsreform, 98*(1), 35–47. https://doi.org/10.1515/mks-2015-980103

Müller-Isberner, R., Gonzales Cabeza, S., & Eucker, S. (2000). *Die Vorhersage sexueller Gewalttaten mit dem SVR-20* [Predicting violent offending with the HCR-20]. Institut für forensische Psychiatrie.

Müller-Isberner, R., Jöckel, D., & Gonzalez Cabeza, S. (1998). *Die Vorhersage von Gewalttaten mit dem HCR-20* [Predicting violent offending with the HCR-20]. Institut für forensische Psychiatrie.

Neal, T. M. S. (2018). Discerning bias in forensic psychological reports in insanity cases. *Behavioral Sciences and the Law, 1–14*. https://doi.org/10.1002/bsl.2346

Neal, T. M. S., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry. *Criminal Justice and Behavior, 41*(12), 1406–1421. https://doi.org/10.1177/0093854814548449

Neal, T. M. S., Slobogin, C., Saks, M. J., Faigman, D. L., & Geisinger, K. F. (2019). Psychological Assessments in Legal Contexts: Are Courts Keeping "Junk Science" Out of the Courtroom? *Psychological Science in the Public Interest*, *20*(3), 135–164. https://doi.org/10.1177/1529100619888860

Nedopil, N. (2005). The impact of base-rates of criminal recidivism on long-term outcome – Results of a follow up study of an unselected group of criminal offenders. In L. A. J. Casselman (Ed.), *Gerechtelijke geestelijke gezondheidszorg: Wetenschap, Beleid en Praktijk* (pp. 107–132). Universitaire Pers Leuven.

Nedopil, N. (2005). *Prognosen in der Forensischen Psychiatrie - ein Handbuch für die Praxis.* [Risk assessment in forensic psychiatry - a practical handbook]. Pabst.

Nicholls, T. L., Pritchard, M. M., Reeves, K. A., & Hilterman, E. (2013). Risk assessment in intimate partner violence: A systematic review of contemporary approaches. *Partner Abuse, 4*(1), 76–168. https://doi.org/10.1891/1946-6560.4.1.76

Nitsche, K., Etzler, S., Balas, J., Eher, R., & Rettenberger, M. (2022). A field study of acute dynamic risk assessment in individuals convicted of sexual offenses. *Psychological Assessment, 34*(6), 583–594. https://doi.org/10.1037/pas0001123

Nowara, S. (1995a). Externe Prognosegutachten im Maßregelvollzug [External risk assessment reports in forensic psychiatry]. *Recht und Psychiatrie*, *13*, 67–72.

Nowara, S. (1995b). *Gefährlichkeitsprognosen bei psychisch kranken Straftätern: Untersuchung zur Qualität der Gutachten gemäss 14 Abs. 3 MRVG NW* [Assessing dangerousness in mentally ill offenders. Examination of the quality of expert witness reports according to article 14 Abs. 3 MRVG NW of the German penal code]. Fink.

Passow, D. (2009). *Zur Qualität forensisch-psychiatrischer Sachverständigengutachten bei Sexualstraftätern mit angeordneter Sicherungsverwahrung [On the quality of forensic-psychiatric expert witness reports about sexual offenders of preventive detention]* [Doctoral dissertation, University Rostock]. E-Publikationen University Rostock. https://doi.org/10.18453/rosdok_id00000666

Pfister, W. (2019, October 11). *Was der Richter vom Sachverständigen erfahren muss [What the judge needs to know from the expert witness]*. 34. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), Munich, Germany.

Prüter-Schwarte, C., Lutz, M., & Kunert, H. J. (2019). Die Qualität von Schuldfähigkeitsgutachten im Maßregelvollzug untergebrachter Patienten - Ergebnisse einer empirischen Studie [The quality of criminal responsibility reports of individuals of forensic psychiatric hospitals - results of an empirical study]. *Der medizinische Sachverständige, 115*(5), 203–207.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk* (2nd ed.). American Psychological Association. https://doi.org/10.1037/11367-000

Rasch, W. (1986). *Forensische Psychiatrie [Forensic Psychiatry]*. Kohlhammer.

Rehder, U., & Suhling, S. (2006). *Rückfallrisiko bei Sexualstraftätern (RRS). Verfahren zur Bestimmung von Rückfallgefahr und Behandlungsnotwendigkeit* [Recidivism risk in sex offenders (RRS). Assessment instrument for recidivism risk and treatment need] (4th ed.). Kriminalpädagogischer Verlag.

Rettenberger, M. (2018). Intuitive, klinisch-idiographische und statistische Kriminalprognosen im Vergleich – Die Überlegenheit wissenschaftlich strukturierten Vorgehens [A conceptual comparison of intuitive, clinical idiographic and actuarial risk assessment approaches – the superiority of science-based methods]. *Forensische Psychiatrie, Psychologie, Kriminologie, 12*(1), 28–36. https://doi.org/10.1007/s11757-017-0463-y

Rettenberger, M. (2019). Instrumente und Methoden zur Einschätzung des Rückfallrisikos [Instruments and methodological approaches for assessing recidivism risk]. In B. Eusterschulte, S. Eucker & P. Born (Eds.), *Forensische Psychiatrie zwischen Wissenschaft und Praxis – Festschrift für Rüdiger Müller-Isberner* (pp. 19–39). Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Rettenberger, M., Briken, P., Turner, D., & Eher, R. (2015). Sexual offender recidivism among a population-based prison sample. *International Journal of Offender Therapy and Comparative Criminology, 59*(4), 424–444. https://doi.org/10.1177/0306624X13516732

Rettenberger, M., & Eher, R. (2012). Möglichkeiten und Grenzen der Verwendung von Prognoseinstrumenten bei Sexualstraftätern [Possibilities and limitations of the use of Risk assessment instruments in sexual offenders]. In J. M. Müller, M. Rösler, P. Briken, P. Fromberger & K. Jordan (Eds.), *EFPPP Jahrbuch 2012–Empirische Forschung in der Forensischen Psychiatrie, Psychologie und Psychotherapie* (pp. 121–126). Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Rettenberger, M., & Eher, R. (2013). SARA – Spousal Assault Risk Assessment Guide. In M. Rettenberger & F. von Franquè (Eds.), *Handbuch kriminalprognostischer Verfahren* (pp. 289–300). Hogrefe.

Rettenberger, M., & Eher, R. (2016). Potenzielle Fehlerquellen bei der Erstellung von Kriminalprognosen, die gutachterliche Kompetenzillusion und mögliche Lösungsansätze für eine bessere Prognosepraxis [Potential sources of bias and error, the illusion of competence of expert witnesses, and some suggestions for the improvement of clinical practice]. *Recht & Psychiatrie, 34(*1), 50–57. https://psychiatrie-verlag.de/product/rettenberger-m-eher-r-potenzielle-fehlerquellen-bei-der-erstellung-von-kriminalprognosen-die-gutachterliche-kompetenzillusion-und-moegliche-loesungsansaetze-fuer-eine-bessere-prognosepraxis/

Rettenberger, M., Eher, R., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., Doyle, M., Folino, J. O., Godoy-Cervera, V., Grann, M., Ho, R. M. Y., Large, M. M., Pham, T. H., Nielsen, L. H., Rebocho, M. F., Reeves, K. A., Ruiter, C. de, Seewald, K., & Singh, J. P. (2017). Kriminalprognosen in der Praxis – die Ergebnisse des International Risk Surveys (IRiS) aus Deutschland [Risk assessment in clinical practice – Results of the International Risk Survey (IRiS) from Germany]. *Diagnostica, 63*(1), 2–14. https://doi.org/10.1026/0012-1924/a000168

Rettenberger, M., Gregório Hertz, P., & Eher, R. (2017). *Die deutsche Version des Violence Risk Appraisal Guide-Revised (VRAG-R)* [The German version of the Violence Risk Appraisal Guide-Revised (VRAG-R)]. Kriminologische Zentralstelle (KrimZ).

Rettenberger, M., Haubner-MacLean, T., & Eher, R. (2013). The contribution of age to the Static-99 risk assessment in a population-based prison sample of sexual offenders. *Criminal Justice and Behavior, 40*(12), 1413–1433.

Rettenberger, M., Rice, M. E., Harris, G. T., & Eher, R. (2017). Actuarial risk assessment of sexual offenders: The psychometric properties of the Sex Offender Risk Appraisal

Guide (SORAG). *Psychological Assessment, 29*(6), 624–638. https://doi.org/10.1037/pas0000390

Rettenberger, M., & von Franqué, F. (2013). *Handbuch kriminalprognostischer Verfahren* [Handbook of risk assessment instruments]. Hogrefe.

Richards, P. M., Geiger, J. A., & Tussey, C. M. (2015). The dirty dozen: 12 Sources of bias in forensic neuropsychology with ways to mitigate. *Psychological Injury and Law, 8*(4), 265–280. https://doi.org/10.1007/s12207-015-9235-1

Rieger, M., & Stadtland, C. (2013). Forensische Psychiatrie [Forensic psychiatry]. In H. Willner, M. Rentrop, R. Müller & J. Bäuml (Eds.), *Klinikleitfaden Psychiatrie Psychotherapie* (5th ed.) (pp. 507–522). Urban & Fischer.

Rice, M. E., & Harris, G. T. (2005). Comparing effect sizes in follow-up studies: ROC area, Cohen's d, and r. *Law and Human Behavior, 29*(5), 615–620. https://doi.org/10.1007/s10979-005-6832-7

Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide—Revised (VRAG-R). *Psychological Assessment, 25*(3), 951–965. https://doi.org/10.1037/a0032878

Riegl, M. (2009). *Die Qualität forensischer Prognosegutachten bei Gewalt- und Sexualstraftätern* [The quality of risk assessment reports about violent and sexual offenders]. GRIN Verlag.

Rorschach, H. (1927). *Rorschach Test – Psychodiagnostic Plates*. Hogrefe.

Rösch, D., Ruckelshaußen, S., Kirsch, J., Gerhards, S., Sroka, L. A., & Imhoff, R. (2020). Schwachsinnig und abartig? Wahrgenommene Stigmatisierung und tatsächliche Etikettierungseffekte der alten und modernisierten Eingangsmerkmale in § 20 Strafgesetzbuch (StGB). [Moronic and abnormal? Perceived stigmatization and actual labelling effects of the old and modernized entry characteristics in § 20 German Penal Code]. *Forensische Psychiatrie, Psychologie, Kriminologie, 15,* 159–168.

https://doi.org/10.1007/s11757-020-00640-x

Rossegger, A., Laubacher, A., Moskvitin, K., Villmar, T., Palermo, G. B., & Endrass, J. (2011). Risk assessment instruments in re-peat offending: The usefulness of FOTRES. *International Journal of Offender Therapy and Comparative Criminology*, *55*(5), 716–731. https://doi.org/10.1177/0306624X09360662

Schläfke, D., Auer, U., König, C., Schnoor, K., Reberning, E., Rachvoll, U., Häßler, F., Keller, F., & Fegert, J. M. (2006). Gutachtenanalyse von Sexualdelikten in Mecklenburg-Vorpommern [Analysis of assessment reports of sexual offenses in Mecklenburg-Vorpommern]. *Forensische Psychiatrie Psychotherapie, 13*(2), 29–58.

Schläfke, D., Lüdemann, W., & Häßler, F. (2000). Qualität der Begutachtung von Sexualstraftätern im Vergleich 1980-1989 vs. 1990-1999 auf der Grundlage eines Dokumentationssystems. [Quality of the assessment of sexual offenders in comparison of the cohorts 1980-1989 vs. 1990-1999 on the basis of a documentation system]. In J. M. Fegert & F. Häßler (Eds.), *Qualität forensischer Begutachtung, insbesondere bei Jugenddelinquenz und Sexualstraftaten* (pp. 257–274). Centaurus.

Schmitt, C., & Rettenberger, M. (2015). Wie (treff-)sicher ist die Einweisung in den Maßregelvollzug? – Forensische, methodische und kriminalpolitische Aspekte [How are the hit rates of placements to forensic psychiatric hospital? - Forensic, methodological, and criminal policy aspects]. *Forensische Psychiatrie und Psychotherapie, 22*(3), 149–169. https://krimpub.krimz.de/frontdoor/deliver/index/docId/392/file/Rettenberger_FPP_2015_3.pdf

Schneider, F., Frister, H., & Olzen, D. (2015). *Begutachtung psychischer Störungen* [Assessment of mental disorders]. Springer. https://doi.org/10.1007/978-3-642-54765-2

Schnoor, K. (2009). *Beurteilung der Schuldfähigkeit. Eine empirische Untersuchung zum*

Umgang der Justiz mit Sachverständigen [Assessing criminal responsibility. An empirical study on the handling of expert witnesses by the judiciary]. Psychiatrie Verlag.

Schöttle, A. (2013). *Die Schuldfähigkeitsbegutachtung in Jugendstrafverfahren: eine Bestands- und Qualitätsanalyse (Bd. 41)* [The criminal responsibility assessment in juvenile criminal proceedings: an analysis of status and quality (Vol. 41)]. LIT Verlag.

Serafim, A. D. P., Saffi, F., Silva, T. G. B. D., Almeida, C. V. D., Hokama, E., Barros, D. M. D., & Dias, A. M. (2015). Forensic neuropsychological assessment: A review of its scope. *Archives of Clinical Psychiatry (São Paulo), 42*(2), 63–67. https://doi.org/10.1590/0101-60830000000049

Singer, J. C., Rettenberger, M., & Boer, D. P. (2016). Further support for a convergent approach to sex offender risk assessment. In L. A. Craig & M. Rettenberger (Eds.), *The Wiley-Blackwell Handbook on the Theories, Assessment, and Treatment of Sexual Offending – Volume II: Assessment* (pp. 695–713). Wiley-Blackwell.

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., & Dean, K. et al. (2014). International perspectives on the practical application of violence risk assessment: survey of 44 countries. *International Journal of Forensic Mental Health, 13*(3), 193–206. https://doi.org/10.1080/14999013.2014.922141

Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and meta-regression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review, 31*(3), 499–513. https://doi.org/10.1016/j.cpr.2010.11.009

Skeem, J. L., & Monahan, J. (2011). Current directions in violence risk assessment. *Current Directions in Psychological Science*, *20*(1), 38–42. https://doi.org/10.1177/0963721410397271

Sklenarova, H., Neutze, J., Kretschmer, T., & Nitschke, J. (2020). Granting leave to patients

in bavarian forensic-psychiatric hospitals: A survey to describe the current process and develop guidelines. *Frontiers in Psychiatry, 11*, 287. https://doi.org/10.3389/fpsyt.2020.00287

Stübner, S., Groß, G., & Nedopil, N. (2006). Inpatient risk management with mentally ill offenders: Results of a survey on clinical decision-making about easing restrictions. *Criminal Behaviour and Mental Health, 16*(2), 11–123. https://doi.org/10.1002/cbm.619

Stübner, S., Steinböck, H., & Fiedler, F. (2018). Schuldfähigkeitsgutachten - Analyse nach Qualitätsmerkmalen (Abstract) [Criminal responsibility assessment reports – analysis of quality aspects (abstract). *Forensische Psychiatrie und Psychotherapie, 25*(3), 326–328.

Suhling, S. (2003). Begutachtung bei Lockerungsentscheidungen [Criminal risk assessment regarding the granting of privileges]. *Neue Kriminalpolitik, 815*(3), 85–88. https://doi.org/10.5771/0934-9200-2003-3-85

Taxman, F. (2017). *Handbook on risk and need assessment: Theory and practice.* Routledge.

Tully, R. J., Chou, S., & Browne, K. D. (2013). A systematic review on the effectiveness of sex offender risk assessment tools in predicting sexual recidivism of adult male sex offenders. *Clinical Psychology Review, 33*(2), 287–316. http://doi.org/10.1016/j.cpr.2012.12.002

Turgut, T., Lagace, D., Izmir, M., & Dursun, S. (2006). Psikiyqtri kliniklerinde s¸iddet ve agresyonun deg˘erlendirilmesi: Tanisal yaklas¸imlar [Assessment of violence and aggression in psychiatric settings: Descriptive approaches]. *Klinik Psikofarmakoloji Bülteni /Bulletin of Clinical Psychopharmacology, 16*(3), 179–194. https://psychiatry-psychopharmacology.com/en/assessment-of-violence-and-aggression-in-psychiatric-settings-descriptive-approaches-161270

Varela, J. G., & Conroy, M. A. (2012). Professional competencies in forensic psychology.

Professional Psychology: Research and Practice, 42, 410–421. http://dx.doi.org/10.1037/a0026776

Verrel, T. (1995). *Schuldfähigkeitsbegutachtung und Strafzumessung bei Tötungsdelikten. Bd. 14: Neue Kriminologische Studien* [Criminal responsibility assessment and sentencing in homicide. Vol. 14: New criminological studies]. Wilhelm Fink Verlag.

Verrel, T. (2015). Schuldfähigkeitsgut- und schlechtachten. Zur Überfälligkeit verbindlicher Qualitätsanforderungen an Gutachten und Urteile [(Bad) Criminal responsibility assessment report. On the overdue nature of binding quality requirements for expert reports and judicial judgments]. In B. Bannenberg, H. Brettel, G. Freund, B.-D. Meler, H. Remschmidt & C. Safferling (Eds.), *Über allem: Menschlichkeit. Festschrift für Dieter Rössner* (pp. 428–442). Nomos.

Viglione, D. J., & Rivera, B. (2003). Assessing personality and psychopathology with projective tests. In J. R. Graham & J. A. Naglieri (Eds.), Comprehensive Handbook of Psychology: *Assessment Psychology*, 10, 531–553. Wiley. https://doi.org/10.1002/0471264385.wei1023

Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment, 17*(3), 377–395. https://doi.org/10.1177/1073191109359587

Viljoen, J. L., Vargen, L. M., Cochrane, D. M., Jonnson, M. R., Goossens, I., & Monjazeb, S. (2021). Do structured risk assessments predict violent, any, and sexual offending better than unstructured judgment? An umbrella review. *Psychology, Public Policy, and Law, 27*(1), 79–97. https://psycnet.apa.org/doi/10.1037/law0000299

von Franqué, F. (2013). Strukturierte, professionelle Risikobeurteilungen [Structured professional risk assessment.]. In M. Rettenberger & F. von Franqué (Eds.), *Handbuch kriminalprognostischer Verfahren* (pp. 357–380). Hogrefe.

Warren, J., Murrie, D., Stejskal, W., Colwell, L., Morris, J., Chauhan, M., & Dietz, P. (2006). Opinion formation in evaluating the adjudicative competence and restorability of criminal defendants: A review of 8,000 evaluations. *Behavioral Sciences and the Law, 24*(2), 113–132. https://doi.org/10.1002/bsl.699

Watkins, C. E., Campbell, V. L., Nieberding, R., & Hallmarlk, R. (1995). Contemporary practice of psychological assessment by clinical psychologists. *Professional Psychology: Research and Practice, 26*(1), 54–60. https://doi.org/10.1037/0735-7028.26.1.54

Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing the Risk for Violence (Version 2).* Simon Fraser University, Health, Law, and Policy Institute.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short Term Assessment of Risk and Treatability (START) Version 1.0 (consultation edition).* St. Joseph's Healthcare; Forensic Psychiatric Services Commission.

Wechsler, D. (1997). WAIS-IV: *Wechsler adult intelligence scale-revised*. Psychological Corporation.

Wertz, M. (2016) *Zur Qualität forensischer Legalprognosegutachten bei Gewalt- und Sexualstraftätern – Eine empirische Validierung* [On the quality of risk assessment reports about violent and sexual offenders – An empirical validation] (master´s thesis, Albert-Ludwigs-University of Freiburg i. Br.].

Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R., & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. *Psychology, Crime, & Law*. Advance online publication. https://doi.org/10.1080/1068316X.2022.2063286

Wertz, M., & Kury, H. (2017a). Verbesserung der Qualität von Prognosegutachten seit der Veröffentlichung von Mindeststandards? Eine empirische Validierung im Zeitverlauf. [Improvements of the quality of criminal risk assessment reports since the publication

of methodological minimum standards? An empirical validation over time]. In J. L. Müller, P. Briken, M. Rösler, M. Müller, D. Turner & W. Retz (Eds.), *EFPPP Jahrbuch 2017 – Empirische Forschung in der Forensischen Psychiatrie, Psychologie und Psychotherapie* (pp. 107–123). Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Wertz, M., & Kury, H. (2017b). Prognosefehler vs. Prognoseirrtümer: Empirischer Zusammenhang zwischen der Einhaltung von Mindestanforderungen und der Trefferquoten von Prognosegutachten laut BZR. [Prognostic failures vs. mistakes: Empirical relation between compliance with methodological minimum requirements and hit rates of risk assessment reports reports according to Federal Central Register] In N. Saimeh (Ed.), *Therapie und Sicherheit im Maßregelvollzug* (pp. 271–287). Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Wertz, M., Kury, H., & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [The application of methodological minimum requirements for risk assessment reports in clinical practice – An empirical validation using officially registered re-offenses]. *Forensische Psychiatrie, Psychologie, Kriminologie, 12*(1), 51–60. https://doi.org/10.1007/s11757-017-0458-8

Wertz, M., Schiltz, K., Imhoff, R., & Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosbegutachtung [The influence of judicial orders on the quality of criminal risk assessment reports]. *Recht & Psychiatrie, 38*(4), 193–200. https://krimpub.krimz.de/frontdoor/deliver/index/docId/335/file/Wertz_RuP_2020_4.pdf

Wertz, M., & Tippelt, S. (2019). Einsatz und Anwendung testpsychologischer Untersuchungen im Rahmen psychiatrischer Gutachten [Use and application of test-

psychological assessment in the context of psychiatric reports]. In F. M. Segmiller & M. Dudeck (Eds.), *Psychiatrische Beispielgutachten: Ein Praxisbuch für Einsteiger und Fortgeschrittene* (1st ed.) (pp. 402–450). Kohlhammer.

Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice, 48*(2), 73–78. https://doi.org/10.1037/pro0000086

Yang, M., Wong, S. C. P., & Coid, J. (2010). The efficacy of violence prediction: A meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin, 136*(5), 740–767. https://doi.org/10.1037/a0020473

Yoon, D., Motekallemi, S., Rettenberger, M., & Briken, P. (2013). Ambulante Behandlung von Sexualstraftätern. Möglichkeiten und Unmöglichkeiten [Outpatient treatment of sexual offenders. Possibilities and impossibilities.]. *Forensische Psychiatrie, Psychotherapie, Kriminologie, 7*(3), 177–182. http://doi.org/10.1007%2Fs11757-013-0220-9

Zapf, P. A., Hubbard, K. L., Cooper, V. G., Wheeles, M. C., & Ronan, K. A. (2004). Have the courts abdicated their responsibility for determination of competency to stand trial to clinicians? *Journal of Forensic Psychology Practice, 4*(1), 27–44. https://doi.org/10.1300/J158v04n01_02

ERKLÄRUNG

gemäß § 6 Absatz 2 g) und gemäß § 6 Absatz 2 h) der Promotionsordnung der Fachbereiche 02, 05, 06, 07, 09 und 10 vom 04. April 2016

Name: _____

Vorname: _____

Hiermit erkläre ich, dass ich die eingereichte Dissertation selbständig, ohne fremde Hilfe verfasst und mit keinen anderen als den darin angegebenen Hilfsmitteln angefertigt habe, dass die wörtlichen oder dem Inhalt nach aus fremden Arbeiten entnommenen Stellen, Zeichnungen, Skizzen, bildlichen Darstellungen und dergleichen als solche genau kenntlich gemacht sind.

Von der Ordnung zur Sicherung guter wissenschaftlicher Praxis in Forschung und Lehre und zum Verfahren zum Umgang mit wissenschaftlichem Fehlverhalten habe ich Kenntnis genommen.

Meine Erklärung bezieht sich auf Schriften, die ich als alleiniger Autor eingereicht habe oder bei Ko-Autorenschaften auf jene Teile, für die ich mich verantwortlich zeichne.

Ich habe keine Hilfe von kommerziellen Promotionsberatern in Anspruch genommen.

_____ _____

Datum Unterschrift

CURRICULUM VITAE

MAXIMILIAN WERTZ

geb. 29. Juli 1988
in Menden (Sauerland)

PROFIL

Psychologe (M.Sc.)

Wissenschaftlicher Mitarbeiter

In Ausbildung zum Fachpsychologen für Rechtspsychologie der DGPs/des BDP

KONTAKT

Abteilung für Forensische Psychiatrie
Nußbaumstraße 7
80336 München

maximilian.wertz@med.uni-muenchen.de

+49 176 57847891

PRAXISERFAHRUNG

Seit 05.2018
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Wissenschaftlicher Mitarbeiter (Prof. Dr. Schiltz)
- an der Abteilung für Forensische Psychiatrie der Klinik und Polklinik für Psychiatrie und Psychotherapie

07.2016 – 06.2018
UNIVERSITÄT REGENSBURG
Wissenschaftlicher Mitarbeiter (Prof. Dr. Osterheider)
- an der Abteilung für Forensische Psychiatrie und Psychotherapie
- Mitarbeit im Dunkelfeldprojekt „Kein Täter werden“

04.2016 – 06.2016
UNIVERSITÄT REGENSBURG
Klinisches Praktikum (Prof. Dr. Osterheider)
- an der Sexualwissenschaftlichen Ambulanz, Projekt „Kein Täter werden“

07.2015 – 09.2015
LUDWIG-MAXIMILIANS-UNIVERSTIÄT MÜNCHEN
Forschungspraktikum (Prof. Dr. Nedopil)
- an der Abteilung für Forensische Psychiatrie

11.2014 – 03.2016
ALBERT-LUDWIGS-UNIVERSITÄT FREIBURG
Wissenschaftliche Hilfskraft (Prof. Dr. Halsband)
- in der Abteilung für Neuropsychologie

07.2014 – 08.2014
LANDESKRIMINALAMT NIEDERSACHSEN
Forschungspraktikum
- in der Kriminologischen Forschungsstelle und Statistik

03.2014 – 03.2014
LWL ZENTRUM FÜR FORENSICHE PSYCHIATRIE LIPPSTADT
Klinisches Praktikum
- mit Schwerpunkt Diagnostik

07.2013 – 09.2013
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Klinisches Praktikum (Prof. Dr. Nedopil)
- an der Abteilung für Forensische Psychiatrie (Diagnostik und psychologische Testdurchführung)

10.2012 – 06.2013
ALPEN-ADRIA-UNIVERSITÄT KLAGENFURT
Forschungspraktikum (Prof. Dr. Renner)
- am Institut für Psychologie (Abteilung für Klinische Psychologie Psychotherapie und Psychoanalyse)

AUSLANDSSEMESTER

2009 – 2010
zwei Auslandssemester an der Universidad Pablo de Olavide, Sevilla, Spanien im Rahmen des Studiengangs INTERNATIONAL BUSINESS STUDIES

STIPENDIEN

Nachwuchsförderpreis 2017
Der Fachgruppe Rechtspsychologie der DGPs/BDP für die beste Masterarbeit

Tagungsstipendium 2016
Für den 50. Kongress der DGPs 2016 der Fachgruppe Rechtspsychologie/Newsletter für den Nachwuchs Rechtspsychologie

Leistungsstipendium 2012/2013
der Alpen-Adria Universität Klagenfurt für hervorragende Leistungen im Studienjahr 2012/2013

Leistungsstipendium 2011/2012
der Alpen-Adria Universität Klagenfurt für hervorragende Leistungen im Studienjahr 2011/2012

ENGAGEMENT

2013
STUDIENASSISTENZ/TUTOR
für den Institutsabteilungsleiter der Abteilung für Angewandte Psychologie und Methodenforschung an der Alpen-Adria Universität Klagenfurt (Prof. Dr. Mayring)

2012 - 2013
STUDIENASSISTENZ/TUTOR
am Institut für Psychologie der Alpen-Adria-Universität Klagenfurt (Prof. Dr. Renner)

AUSBILDUNG

Seit 10.2017
JOHANNES GUTENBERG-UNIVERSITÄT MAINZ
Promovend (Prof. Dr. Rettenberger)
- an der Abteilung Sozial- u. Rechtspsychologie

Seit 2016
IN WEITERBILDUNG ZUM FACHPSYCHOLOGEN FÜR RECHTSPSYCHOLOGIE DER DGPs/BDP

Seit 2016
IN WEITERBILDUNG ZUM SEXUALTHERAPEUTEN DER AVM Salzburg

10.2014 – 09.2016
KLINISCHE PSYCHOLOGIE, (M.Sc.)
Albert-Ludwigs-Universität Freiburg
- Schwerpunkt: Neuro- und Rehabilitationswissenschaften
- Akadem. LehrerInnen: Prof. Dr. Bengel, Prof. Dr. Halsband, Prof. Dr. Heinrichs, Prof. Dr. Prof. Tuschen-Caffier
- Masterthesis: Zur Qualität forensischer Legal-prognosegutachten bei Gewalt- und Sexualstraftätern - Eine empirische Validierung (Beurteilung: 1,0; Prof. Dr. Halsband)
- Abschlussnote: 1,3

10.2012 – 09.2015
ANGEWANDTE BETRIEBSWIRTSCHAFTSLEHRE (M.Sc.)
Alpen-Adria Universität Klagenfurt
- Schwerpunkt: Personalführung
- Masterthesis: Mitarbeiterführung an deutschsprachigen Universitäten – Empirische Überprüfung der Korrelate narzisstischer Führung (Beurteilung: 1,0)

10.2011 – 09.2013
PSYCHOLOGIE (B.Sc.)
Alpen-Adria Universität Klagenfurt
- Schwerpunkt: Klinische Psychologie
- Akadem. LehrerInnen: Prof. Dr. Mayring, Prof. Dr. Renner, Prof. Dr. Vitouch, Prof. Dr. Glück, Prof. Dr. Alexandrowicz
- Bachelorthesis: Affektive Länderprofile Deutschlands und Österreichs (Beurteilung: 1,0; Prof. Dr. Renner)
- Abschlussnote: 1,1

10.2007 – 09.2011
INTERNATIONAL BUSINESS STUDIES (B.A.)
Universität Paderborn
- Schwerpunkt: Arbeits- und Organisationspsychologie / Personalauswahl / Recruiting

07.1999 – 06.2007
ABITUR
Walburgis-Gymnasium, Menden
- 7. Klasse übersprungen

SPRACHEN

Deutsch	●●●●●
Englisch	●●●●○
Spanisch	●●●●○
Französisch	●●●○○

ZERTIFIKATE

Institut für Gewaltforschung u. Prävention (IGF)

Schulung der PCL-R und PCL-SV (Basiskurs), Prof. Dr. Mokros

Risikoeinschätzung und Messung der risikorelevanten Veränderung mittels der VRS und VRS:SO, Prof. Dr. Eher

Kriminologische Zentralstelle (KrimZ)

Workshop zur praktischen Anwendung und Interpretation des VRAG-R (Zertifizierungskurs), Prof. Dr. Rettenberger

Workshop zur praktischen Anwendung und Interpretation von Static-99, Stable-2007 und Acute-2007 (Zertifizierungskurs), Prof. Dr. Rettenberger

UNIVERSITÄRE LEHRTÄTIGKEITEN

Seit 05.2018

Mitgestaltung der Vorlesung „Forensische Psychiatrie“ für Studierende der Medizin, Rechtswissenschaften und Psychologie an der LMU München

Eigenständige Gestaltung von Studentenkursen in der Seminarreihe „Bedside Teaching“ an der LMU München

BETREUUNG VON ABSCHLUSSARBEITEN

„Zwei Jahre *Kein Täter Werden* am Standort München - Projekt- und Patientenbezogene Merkmale unter besonderer Berücksichtigung der Unterschiede zwischen Hell- und Dunkeldpatienten“ (Masterarbeit; Nele Brenzinger, Ludwig-Maximilians-Universität München, in preparation)

„Forensische Profile strafrechtlicher Begutachtungsprobanden mit Schizophrenie, schizotypen und wahnhaften Störungen“ (Masterarbeit; Lea Wassermann, Universität Regensburg, in preparation)

„Forensische Profile strafrechtlicher BegutachtungsprobandInnen mit affektiven Störungen – eine retrospektive Analyse“ (Bachelorarbeit; Sina Wessels, Deutsche Hochschule für Gesundheit & Sport, 2022)

„Pathological and non-pathological personality of female forensic psychiatric patients: Exploring the relationship to crime and demographics“ (Masterarbeit; Ana Macchia, Universität Konstanz, 2021)

„Profiles of Female Offenders – Biography, Personality, and Criminal Behaviour of Forensic Examinees“ (Masterarbeit; Caroline Blunck, Maastricht University, 2021)

„Relevance of Offender-Related Characteristics in Criminal Risk Assessment of Sexual Offenders“ (Masterarbeit; Maren Giersiepen, Heinrich-Heine-Universität Düsseldorf, 2021)

„Persönlichkeitsprofile von Maßregelvollzugspatientinnnen. Explorative, clusteranalytische Persönlichkeitsbestimmung von Extraversion, Neurotizismus, Verträglichkeit, Psychopathie und Empathie“ (Masterarbeit; Julia Thaler, Universität Regensburg, 2021)

„Vergleich der Persönlichkeitsprofile nach §§ 63, 64 StGB untergebrachter Maßregelvollzugspatientinnen und inhaftierter Straftäterinnen unter besonderer Berücksichtigung von Aggressionsbereitschaft und Wut/Ärger-Ausdruck (FAF, STAXI-2, PFI+)“ (Masterarbeit; Lara Heck, Eberhard-Karls-Universität Tübingen, 2021)

„Forensische Profile strafrechtlich begutachteter Kapitaldelinquenter. Biografie, Persönlichkeit und psychische Störungsbilder“ (Masterarbeit; Vivian Schmid, Julius-Maximilians-Universität Würzburg, 2021)

„Profile forensisch begutachteter StalkerInnen. Persönlichkeit, Biographie, Typologie & Vorgehen“ (Masterarbeit; Timon Rotter, Ludwig-Maximilians-Universität München, 2021)

„Profile strafrechtlich begutachteter Kindesmissbrauchs- und Kinderpornographiedelikttäter: Biografie, Modus Operandi und Typologie“ (Masterarbeit; Lina Willfurth, Ludwig-Maximilians-Universität München, 2021)

„Merkmale psychiatrisch begutachteter, aber nicht psychisch kranker Rechtsbrecher“ (Masterarbeit; Korbinian Gailer, Ludwig-Maximilians-Universität München, 2021)

„Forensische Profile persönlichkeits- und sexualpräferenzgestörter strafrechtlicher Begutachtungsprobanden“ (Masterarbeit; Sandra von der Heyden, SRH Hochschule Heidelberg, 2021)

„Kognitive Verzerrungen und emotionale Überidentifikation bei sexuellen Kindesmissbrauchern – Normnierung der KV-M und EKK-R anhand einer strafrechtlichen Begutachtungsstichprobe“ (Bachelorarbeit; Lisa Baumüller, Universität Regensburg, 2021)

„Forensische Profile strafrechtlich begutachteter Brandstifter – Biographie, Persönlichkeit, Deliktmerkmale und Typologie“ (Bachelorarbeit; Lynn Osselmann, Ludwig-Maximilians-Universität München, 2021)

„Einsatz und Bedeutung testpsychologischer Verfahren im Rahmen psychiatrischer Begutachtungen zur Kriminalprognose (Masterarbeit; Lisa Hank, Universität Regensburg, 2020)

„Empirische Validierung der Umsetzung der Mindestanforderungen für Schuldfähigkeitsgutachten in der Praxis“ (Masterarbeit; Nicolo Salvo, Hochschule Fresenius, 2020)

„Persönlichkeitsprofile inhaftierter StraftäterInnen unter besonderer Berücksichtigung von Emotionserleben und Empathie“ (Masterarbeit; Annika Bäumler, Hochschule Fresenius, 2020)

Persönlichkeitsprofile nach §§ 63, 64 StGB untergebrachter Maßregelvollzugspatientinnen und inhaftierter Straftäterinnen unter besonderer Berücksichtigung von Konfliktbewältigung und Frustrationstoleranz“ (Masterarbeit; Maria Schöpp, Hochschule Fresenius, 2020)

„Einsatz und Bedeutung testpsychologischer Verfahren im Rahmen psychiatrischer Begutachtungen zur Schuldfähigkeit“ (Bachelorarbeit; Nina Heiß, Hochschule Fresenius, 2020)

„Aggressionen bei Maßregelvollzugspatientinnen (gem. Unterbringung nach §§ 63/64 StGB“ (Masterarbeit; Anna Schwanghart, Hochschule Fresenius, 2019)

PUBLIKATIONSVERZEICHNIS

ORIGINALARBEITEN / BUCHBEITRÄGE

Popovic, D., **Wertz, M.,** Geisler, C., Kaufmann, J., Lähteenvuo, M., Lieslehto, J., Witzel, J., Bogerts, B., Walter, M., Koutsouleris, N. & Schiltz, K. (in press). Patterns of risk - Using machine learning and structural neuroimaging to identify pedophilic offenders. *Frontiers in Psychiatry.*

Wertz, M., Schobel, S., Schiltz, K., Imhoff, R. & Rettenberger, M. (in press). A comparison of the predictive accuracy of structured and unstructured risk assessment methods. *Psychological Assessment.*

2022

Wertz, M., Giersiepen, M., Schiltz, K., Imhoff, R. & Rettenberger, M. (2022). Consideration and relevance of empirically unsupported risk factors in expert witnesses‘ risk assessment reports (Manuscript submitted for publication). Department of Forensic Psychiatry, Ludwig-Maximilians-University, Munich.

Wertz, M., Schobel, S., Nedopil, N. & Schiltz, K. (2022). Rechtsbereichsspezifische Unterschiede untersuchungs- und probandenbezogener Merkmale innerhalb der forensischen Begutachtungspraxis – Empirischer Überblick am psychiatrischen Universitätsklinikum der Ludwig-Maximilians-Universität München [Forensic assessment report practice at the psychiatric university hospital of the Ludwig-Maximilians-University, Munich – Empirical overview of report- and offender-related variables] (Manuscript submitted for publication). Department of Forensic Psychiatry, Ludwig-Maximilians-University, Munich.

Schiltz, K., Langgartner, S. & **Wertz, M.** (2022). Die Sicherungsverwahrung [Preventive detention]. In: F. Häßler, N. Nedopil & M. Dudeck (Eds.): *Praxishandbuch Forensische Psychiatrie (*p. 660-669). Berlin: Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Wertz, M., Hank, L., Hausam, J., Konrad, N., Schiltz, K. Imhoff, R. & Rettenberger, M. (2022). The use and reporting practice of psychological tests in German risk and criminal responsibility expert reports. *Psychology, Crime & Law*. Advance online publication. https://doi.org/10.1080/1068316X.2022.2063286

2021

Wertz, M.*, Hausam, J.*, Konrad, N., Schiltz, K., Imhoff, R. & Rettenberger, M. (2021). Qualität von Schuldfähigkeitsgutachten – Mindestanforderungen, unterbringungsrelevante Gefährlichkeitsprognose und Berücksichtigung im richterlichen Urteil [Quality of criminal responsibility reports – Minimum requirements, risk assessment, and consideration in court decisions]. Recht & Psychiatrie, 39(4), 202 – 211.
*geteilte Erstautorenschaft

Wertz, M., Mader, E., Nedopil, N., Schiltz, K. & Yundina, E. (2021). Antwortverzerrung oder Symptombelastung? Beschwerdeschilderung von psychiatrischen Patienten und sozialmedizinischen Begutachtungsprobanden [Response distortion or symptom severity? Symptom description by psychiatric patients and sociomedical assessment subjects]. *Der Nervenarzt, 92,* 1163–1171.

Wertz, M. & Rettenberger, M. (2021). Die Verwendung standardisierter Prognoseinstrumente in der Begutachtungspraxis: Empirische Erkenntnisse zur Häufigkeit und Risikokommunikation in Abhängigkeit gutachten- und probandenbezogener Merkmale [The use of standardized risk assessment instruments in the practice of risk assessment: Empirical findings on frequency and risk communication as a function of assessment- and subject-related characteristics.]. *Forensische Psychiatrie und Psychotherapie, 28*(3), 241-261.

2020

Köthe, N., **Wertz, M**., Steffanowski, A. & Yundina, E. (2020). Der Rosenzweig Picture-Frustration Test (PFT) im forensischen Kontext [The Rosenzweig Picture-Frustration Test in forensic psychiatric assessments – a validation study]. Eine Validierungsstudie. *Recht & Psychiatrie, 38*(2), 84-92.

Wertz, M., Schiltz, K., Imhoff, R. & Rettenberger, M. (2020). Der Einfluss des richterlichen Auftrags auf die Qualität der Arbeit von Sachverständigen im Rahmen der Prognosebegutachtung [The influence of the judicial order on the quality of the work of expert witnesses in the context of risk assessment]. *Recht & Psychiatrie, 38*(4), 193 – 200.

Wertz, M., Schobel, S. & Schiltz, K. (2020). Interdisziplinäre Zusammenarbeit in der Forensischen Psychiatrie. *Zeitschrift für Didaktik der Rechtswissenschaft, 7*(2), 151-159.

Wuttke, E., **Wertz, M**., Schiltz, K., A. & Nedopil, N. (2020). Resozialisierungsempfehlungen in Prognosegutachten und deren Umsetzung – Ein Werkstattbericht [Criminal risk assessment and reality – Recommendations for cocial rehabilitation in penal institutions and forensic psychiatric hospitals from a sociological perspective. *Recht & Psychiatrie, 38*(4), 201-209.

2018

Wertz, M., Kury, H. & Rettenberger, M. (2018). Umsetzung von Mindestanforderungen für Prognosegutachten in der Praxis – Eine empirische Validierung unter Berücksichtigung der Rückfallquoten [Implementation of the minimum requirements for risk assessment reports in practice – Empirical validation considering recidivism rates]. *Forensische Psychiatrie, Psychologie, Kriminologie, 12*(1), 51 – 60.

Wertz, M. & Tippelt, S. (2018). Einsatz und Anwendung testpsychologischer Untersuchungen im Rahmen psychiatrischer Gutachten [Use of psychological test examinations in psychiatric assessment reports]. In: F. Segmiller & M. Dudeck (Eds.): *Psychiatrische Beispielgutachten. Ein Praxisbuch für Einsteiger und Fortgeschrittene* (p. 402-450). Stuttgart: Kohlhammer.

2017

Wertz, M. & Kury, H. (2017a). Verbesserung der Qualität von Prognosegutachten seit der Veröffentlichung von Mindeststandards? Eine empirische Validierung im Zeitverlauf. In: J.L. Müller, P. Briken, M. Rösler, M. Müller, D. Turner, W. Retz (Eds.): *EFPPP Jahrbuch 2017 – Empirische Forschung in der Forensischen Psychiatrie, Psychologie und Psychotherapie* (p. 107-123). Berlin: Medizinisch-Wissenschaftliche Verlagsgesellschaft.

Wertz, M. & Kury, H. (2017b). Prognosefehler vs. Prognoseirrtümer: Empirischer Zusammenhang zwischen der Einhaltung von Mindestanforderungen und der Trefferquote von Prognosegutachten laut BZR? In: N. Saimeh (Eds.): *Therapie und Sicherheit im Maßregelvollzug* (p. 271-287), Berlin: Medizinisch-Wissenschaftliche Verlagsgesellschaft.

2015

Renner, W. & **Wertz, M.** (2015). Valence and Efficacy: The Affective meanings of Human Values and their Relationship to Moral Decisions. *International Journal of Business and Social Research, Vol. 5*, 44 – 55.

2014

Renner, W., **Wertz, M**., Fritzsche, S. & Gula, B. (2014). Asymmetric mutual social perception of Austrians and Germans: A social identity approach assessing implicit and explicit attitudes. *International Journal of Business and Social Research, Vol. 4*, 50 – 57.

WISSENSCHAFTLICHE VORTRÄGE

2022

Wertz, M. (2022, September). *Officially registered reoffenses of violent and sexual offenders – Importance of minimum requirements and different methodological approaches of risk assessment reports in clinical practice.* Vortrag auf dem 52. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Hildesheim, Deutschland.

2020

Wertz, M. (2020, Oktober). *Persönlichkeitsprofile forensischer Populationen - Projektstatus und erste Ergebnisse*. Vortrag im Rahmen der 35. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), München.

Wertz, M. (2020, September). *Aspekte der Beschwerdevalidierung mit Fragebogen und neuropsychologischen Tests.* Vortrag im Rahmen des Seminars „Medizinische Begutachtung" der Bayerischen Landesärztekammer, München.

2019

Wertz, M. (2019, Oktober). *Empirische Untersuchungen über die Qualität forensischer Prognoseerstellung – Mindestanforderungen, Risikokommunikation, Testpsychologie.* Vortrag im Rahmen der 34. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), München.

Wertz, M. (2019, September). *Empirische Untersuchungen über die Qualität forensischer Prognoseerstellung – Mindestanforderungen, Risikokommunikation, Testpsychologie.* Vortrag im Rahmen der 18. Tagung der Fachgruppe Rechtspsychologie der Deutschen Gesellschaft für Psychologie, Jena.

Wertz, M. (2019, Juni). *Aspekte der Beschwerdevalidierung mit Fragebogen und neuropsychologischen Tests.* Vortrag im Rahmen des Seminars „Medizinische Begutachtung" der Bayerischen Landesärztekammer, München.

2018

Wertz, M. (2018, Oktober). *Qualität forensischer Prognoseerstellung – Empirische Untersuchungen über die Bedeutung der Mindestanforderungen für Prognosegutachten und unterschiedlicher Ansätze der Prognosemethodik.* Vortrag im Rahmen der 33. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), München.

Wertz, M. (2018, September). *Officially registered reoffenses of violent and sexual offenders – Importance of minimum requirements and different methodological approaches of risk assessment reports in clinical practice.* Vortrag auf dem 51. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Frankfurt a.M., Deutschland.

2017

Wertz, M. (2017, September). *Umsetzung der Mindestanforderungen an Prognosegutachten in der Praxis? Eine empirische Validierung unter Berücksichtigung der Rückfallquoten.* Vortrag im Rahmen der 17. Tagung der Fachgruppe Rechtspsychologie der Deutschen Gesellschaft für Psychologie, Jena.

Wertz, M. (2017, März). *Verbesserung der Qualität von Prognosegutachten seit der Veröffentlichung von Mindeststandards? Eine empirische Validierung im Zeitverlauf unter Berücksichtigung der Rückfallquoten.* Vortrag im Rahmen der 32. Eickelborner Fachtagung zu Fragen der Forensischen Psychiatrie, Lippstadt-Eickelborn.

2016

Wertz, M. (2016, November). *Qualität von Prognosegutachten von Sexual- und Gewaltstraftätern. Eine empirische Validierung.* Vortrag im Rahmen des 6. Symposiums Empirische Forschung in der forensischen Psychiatrie, Psychologie und Psychotherapie (EFPPP), Mainz.

Wertz, M. (2016, Oktober). *Qualität von Prognosegutachten von Sexual- und Gewaltstraftätern. Eine empirische Validierung.* Vortrag im Rahmen der 31. Münchner Herbsttagung der Arbeitsgemeinschaft für Methodik und Dokumentation in der Forensischen Psychiatrie (AGFP), München.

Wertz, M. (2016, Juni). *Qualität von Prognosegutachten von Sexual- und Gewaltstraftätern. Eine empirische Validierung*. Vortrag im Rahmen der 12. Summer Conference, Regensburg.