https://doi.org/10.36315/2026inpact050
Psychological Applications and Trends 2026

QUALITY IN EXPERT WITNESS REPORTS – PEER REVIEW AND COMPUTERIZED EVALUATION: A WAY FORWARD?

Martina Feldhammer-Kahr[1,2], Johannes Woschizka[1], Nina Kaiser[2], Markus Sommer[1], & Martin Arendasy[1]
[1]*Department of Psychology, University of Graz (Austria)*
[2]*Department of Criminal Law, Criminal Procedure Law and Criminology, University of Graz (Austria)*

Abstract

Under Austrian Criminal Law, courts are required to appoint expert witnesses whenever special knowledge is required to conduct investigations or to take evidence, and such expertise is not available within the court or judicial institutions. These experts must possess special expertise and be impartial, whereby the required expertise may derive from a variety of disciplines depending on the needs of the particular case. However, in proceedings concerning the placement of an individual in a forensic therapeutic institution, it primarily has to be a psychiatric expert, preferably one who is also registered in the field of psychiatric criminal prognosis or if a psychiatric expert is not available or cannot be appointed in due time, an expert in clinical psychology may be designated instead. If – in the court's opinion - the expert witness report is indeterminate, or the opinion, contradictory or otherwise flawed or the statements of two experts differ significantly, the court must appoint a further expert if the doubts cannot be removed by questioning. However, if the court needs an expert witness because it lacks the expertise in forensic psychology, how should the court have the expertise to assess the quality, validity and thus the evidential value of expert witness reports – let alone decide between the opinions of two or more disagreeing experts? For many years – especially due to the limited availability of qualified experts and high case numbers - there is an ongoing discussion about the quality of the expert witness reports in these cases and the standards they should meet. Although there is plenty of literature in this field it still seems to be unclear for legal and forensic professionals, which criteria should be met and how this should be evaluated in daily court practice. Kannegießer et al. (2021) seem to show us a way. While this study does not pertain directly to criminal law issues, it nonetheless offers points for discussion, especially given that the fundamental questions of quality standards and their evaluation by "non-experts" apply equally to all branches of law. They propose a peer-review procedure in family related matters. However, fewer expert witnesses are available than courts and other state authorities require, compounded by the additional issue of escalating costs that would come along with a peer-review procedure. A long-term cost-effective and less time-consuming alternative could be a computerized program. The challenges of the construction of such a program will be discussed in this paper.

Keywords: *Quality, expert witness reports, peer-review, computerized program.*

1. Introduction

In some countries like Austria there is a paradox situation considering the Code of Criminal Procedure (e.g., § 126 Austrian Code of Criminal Procedure) and expert witness reports. Judges and prosecutors are required to rely on expert witnesses, when the court lacks the special expertise which is needed e.g., to assess the mental health condition of the accused at the time of the crime and the risk of future criminal behavior. However, the subsequent legal text demands that if an expert witness report is indeterminate, contradictory or otherwise flawed or two expert witnesses do not agree in essential facts or conclusions drawn from the facts, and it is not possible to dispel the courts doubts after an oral interview within the main proceedings, one must fall back on another expert witness (§ 127 (3) Austrian Code of Criminal Procedure). Meaning, the court needs to evaluate the quality of the expert witness reports despite knowing that they do not have the expertise to do so. If they had, they would not need the expert witness in the first place.

Let us keep in mind, psychological and psychiatric experts already face the challenge that the questions they are asked to answer involve issues that are far more complex than purely scientific questions

or hypothesis testing. As Schoemig and Vogel (2025) note, these are questions that concern complex situations and circumstances, in which there can be a multitude of (unpredictable) interactions between the observable variables. Often, the question arises as to whether assumptions about causal relationships can be made in such a way that they are applicable to the individual case. Therefore, it is already very challenging for experts to evaluate the quality of expert reports, unless we are dealing with extreme cases of very low quality (see Kastner, 2019). How can it be expected of judges to accomplish this, although they are the ones who make the decision?

There are a multitude of areas in which experts provide reports for courts, ranging from family law issues to questions of culpability, recidivism, and much more. Each area has a different approach to assessment and the procedures that should theoretically be applied. For instance, in criminal recidivism prediction, standardized assessment tools can offer useful guidance, potentially even more than behavioral observation. In contrast, in family psychological evaluations, interaction observation between parents and children is a very important component. Each of these areas is dealing with questions of varying complexity. Here, an exemplary research question is presented for illustrative purposes.

A 28-year-old man, was convicted of extortionate kidnapping, aggravated extortion, and grievous bodily harm, and sentenced to five years of imprisonment. At the time of the offense, he was heavily intoxicated, which significantly impaired his ability to control his actions. He has an extensive criminal record, including convictions for drunk driving, property damage, fraud, theft, and resisting police officers, indicating a pattern of repeated misconduct. After three years of imprisonment a forensic psychological risk assessment was conducted to evaluate the likelihood of future offenses, their nature, frequency, and severity, as well as measures to mitigate the risk (Kobbé, 2021).

Different stages of proceedings, initial case situations, and specific questions at hand call for different approaches, and therefore the requirements placed on assessments can only provide a rough framework. A minimum threshold that does not yet reveal anything specific about the real quality of the assessment. But let us briefly examine the minimum requirements for assessments provided by the literature, which are still only partially implemented.

2. Minimum requirements of expert witness reports

The Diagnostic and Test Commission (DTK) of the Federation of German Psychologists' Associations published a list of quality standards and their detailed specifications in 2017, including the (1) specification of the questions and (2) derivation of psychological questions, (3) the planning and justification of information gathering using high-quality and appropriate psychological methods, (4) the justification and (pre-)determination of decision-making strategies to be applied during the assessment, (5) conducting the examination, (6) evaluation of the examination results, (7) transparent, detailed, and accurate presentation of the findings, taking into account the reliability and validity of the methods used, (8) derivation of conclusions from the findings and finally, (9) answering the psychological questions and addressing the questions posed by the commissioning party (Diagnostic and Test Commission, 2017, pp. 2-7).

Westhoff and Kluck (2014) provide guidelines for the evaluation of psychological assessments by non-experts. They guide readers to ask questions about each point that should be included or implemented in an assessment. Two examples of these fifty-one recommendations should give a small insight in the process.

Example 1: “Are the pieces of information weighted according to their significance for the question at hand?” Information can vary in its relevance depending on the quality of the source, its scope, or the way it was obtained. Therefore, the assessor weights the information based on its significance for answering the question (Westhoff & Kluck, 2014, p. 252).

Example 2: “Is the question being answered?” Clients commissioning a psychological assessment seek decision-making support in situations where, in their view, their own psychological expertise is insufficient. Therefore, the client's question must be addressed in the assessment. For example, in legal cases, this question is formulated in the "evidence order." Logically, the question can only be answered once all relevant information has been presented, weighted, and combined (Westhoff & Kluck, 2014, p. 252).

Moreover, a group of experts from the fields of law - including judges of the Federal Court of Justice - as well as experts in forensic psychiatry and forensic psychology, collaborated to develop minimum standards for expert reports addressing questions of culpability risk assessment reports. These minimum standards were designed to help experts produce high-quality reports tailored to the specific needs of criminal proceedings. They were explicitly intended to assist judges in evaluating the quality of these reports. These standards cover: (1) choice of assessment methods, (2) classification systems, (3) extent of psychological disorder, (4) traceability and transparency and (5) evidentiary basis of the report, including

social and biographical characteristics, with particular attention to the temporal consistency of psychopathological findings (Boetticher et al., 2007).

3. Peer-review opportunities

But how can we achieve an improvement in the quality of assessments? Even though there are still issues with the quality of expert reports, a very positive side effect of the discussion around quality criteria is that some authors report an improvement in quality (cf. Fegert et al., 2024).

Fundamentally, questions addressed by experts are often associated with serious consequences resulting from the decisions made by the clients (cf. Kannegießer et al., 2021). Therefore, we should not be satisfied with an improvement trend in only a few individual assessments. Hence, solutions should be found that can be implemented quickly. One proposal is a publicly funded, moderated peer review process for expert assessments (Banse, 2017). Based on this proposal, Kannegießer et al. conducted a pilot project on peer review processes to improve expert reports for family psychological assessments. Fifty-one reviewers of family psychological expert reports submitted their own reports via an online platform and, in exchange, reviewed two other reports based on a standardized evaluation scheme. The expert reports received predominantly positive evaluations (Kannegießer, Wegmann, & Ebner, 2020). As the authors point out, the pilot project aimed to provide peer-to-peer feedback on an equal footing, involving colleagues with comparable qualifications and professional experience. Expert witnesses were able to volunteer and decide which reports to upload. The most significant problem in this context is self-selection (cf. Hu, Pavlou & Zhang, 2017). It can be assumed that those who participated were experts who were already motivated to improve and to critically reflect on their own work. It is rather unlikely that negative cases, such as those described by Kastner (2019), would come forward for such projects.

Belke et al. (2025) focused in their peer-review project on the statement validity assessment. The same problems are also evident in the study. In this case, only 30 experts could be recruited, who provided 33 expert reports. The reports ranged from 18 to 136 pages in length, which illustrates how substantial the effort required from the participants is. Here too, the expert reports were consistently evaluated positively, with little variance in the responses. This may be attributable to the selection process discussed above. On the one hand, experts who carry out their work less conscientiously are unlikely to participate, and on the other hand, participants may selectively choose their own reports to submit (cf. Kannegießer et al., 2020).

Such projects represent initial approaches. As part of the study, the experts also received direct feedback to help improve their work, which in principle has potential. However, just as refinements in training require time to contribute to a general improvement in the quality of expert reports, smaller projects also have the disadvantage that it may take too long to bring about actual change. As mentioned at the outset, expert evaluations involve decisions with far-reaching consequences (e.g., a child unjustifiably not seeing a parent for a period of time; an (false negative) offender being released too early and becoming a recidivist; or a (false positive) person remaining unjustly in a forensic-therapeutic institution).

The questions governments must address are whether expert witnesses can be required to undergo at least occasional peer evaluation once they are already listed as certified experts, and whether such evaluations can be linked to consequences. An incident in Austria prompted a re-evaluation of weapons-psychological assessments. As a result, the Federal Ministry of the Interior (BMI) recently pre-published an amendment to the Ordinance of the Federal Minister of the Interior, which modifies the 1st and 2nd Weapons Act Implementation Ordinances. This amendment introduces measures to enhance the quality of clinical psychological expert reports, including the mandatory use of exploratory interviews and the use of at least three standardized assessment instruments reflecting the current state of scientific research. Experts will furthermore be required to complete comprehensive field-specific training covering (1) legal foundations, particularly in the field of weapons law, (2) discipline-specific clinical-psychological aspects, (3) framework conditions for the preparation of clinical-psychological expert reports, (4) in-depth exploration of the content through cases. A list of registered experts is then maintained by the BMI. Registration will be limited to five years, after which recertification for the next five years depends on the proof of continuing advanced training, as well as completed supervision (Ministry of the Interior, 2026).

This raises the question of whether - in the case of much more complex issues (e.g. recidivism risk assessment), where there is already a shortage of expert witnesses (cf. Kastner, 2019) - it is feasible to place even greater pressure on experts. Computer-based screening tools may offer a potential solution.

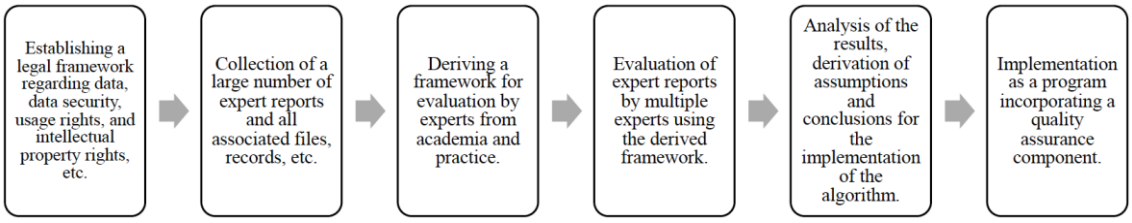
4. Technology-assisted review

One objective would be to develop computer programs that help assess the quality of expert reports. A program that could serve, on the one hand, as a tool for self-evaluation - similar to programs (e.g., TextGuard AI) that help students identify plagiarized passages in their work - and, on the other hand,

as a tool for legal professionals to evaluate expert reports. One possible approach would be the implementation of so-called technology-assisted reviews to determine the relevance of an expert report and the extent to which it serves as a decision-making aid.

To enable a computerized review of expert reports, numerous preliminary steps and a substantial investment of time are required (see concept in Figure 1). First of all, a large number of expert reports (including all relevant case files) would need to be made available and evaluated by experts based on various criteria. In addition, particularly in the context of risk assessment reports in criminal proceedings, it would be desirable to know whether the expert's assessment was accurate (while being aware of the false-positive dilemma) and to what extent the report was considered relevant. Legal frameworks regarding data protection are necessary to safeguard all mentioned or affected individuals. Once an overview of the data situation has been successfully established, experts from academia and practice should develop a framework for evaluating the expert reports (e.g., evaluation of the methods used, etc.). The next step is the training of experts who evaluate the reports (i.e., how to use this framework). Ideally, this framework would be applicable to the evaluation of all expert reports; however, as noted in the introduction, reports prepared for courts are consistently complex. Therefore, an attempt is made to address this problem by having a smaller number of expert reports evaluated in a preliminary study. Once the evaluations are completed, the data analysis focuses on deriving assumptions and conclusions that can later be implemented as a program. Due to the sensitivity of the data and the research questions, the computer program should be largely independent of conventional programs that utilize large language models (e.g. GPT Series). An important aspect is also ensuring quality through a quality assurance component, fundamentally comparable to those used in automatic item generation (Arendasy, 2004). In the case of assessments, it is crucial that the methods used are applicable to individual cases. This means that one cannot simply assume that what was valid for person X must also necessarily apply to person Y, as we are dealing with a multitude of variables. The question is: Can I ensure that a computer program is capable of evaluating not only the nomothetic assessment but also the necessarily idiographic one?

Figure 1. Process flow for the development of a support tool for quality assurance and improvement of assessments.



Furthermore, some fundamental problems arise e.g. like how can I make the quality of expert reports measurable, given that it is not a physical quantity in the actual sense (Schoemig & Vogel, 2025). Bias may be incorporated into automated assessment programs because they are developed by expert clinical judges who themselves may hold biases and may inadvertently embed these biases into the systems they design (Lichtenberger, 2006).

5. Conclusion

Fortunately, there is a broad discourse on the quality of assessments, which, as mentioned earlier, has led to various projects and improvements in the quality of expert reports (cf. Fegert et al., 2024). This is particularly important as the decisions made based on assessments are associated with serious consequences. As outlined above, the implementation of computerized support programs for experts and legal professionals is not straightforward and requires time.

Therefore, an important issue is the education and professional training of forensic psychology experts, alongside training opportunities and guidelines for legal professionals in assessing the quality of forensic psychological reports (cf. Tadei et al., 2016). Projects that enable peer-review for expert witness reports should continue to be promoted, as all involved stakeholders can benefit from the learning process. We should pursue the learning and interdisciplinary exchange, regardless of how quickly the implementation of this computer-assisted aid can be achieved.

References

Arendasy, M. (2004). *Automated Item Generation and Psychometric Quality Assurance Using the Example of the Matrix Test GEOM* (Unpublished habilitation thesis, University of Vienna).

Banse, R. (2017). Qualitätssicherung von rechtspsychologischen Sachverständigengutachten durch ein moderiertes Peer-Reviewverfahren: Ein Vorschlag zur Diskussion [Quality control of legal psychological expert evaluations by a modified peer-review procedure: a proposal for discussion]. *Praxis der Rechtspsychologie*, *27*(2), 113–130.

Belke, A.-P., Pfundmair, M., Wolf-Brandstetter, C., Wegmann, U., & Kannegießer, A. (2025). Abschlussbericht zum Pilotprojekt 'Professionelle Selbstkontrolle. *Online-Peer-Review-Verfahren in der aussagepsychologischen Begutachtung' - Kurzfassung*. [Final report on the pilot project 'Professional self-monitoring: online peer-review procedure in forensic psychological assessment' – a summary.]

Boetticher, A., Nedophil, N., Bosinski, H. A. G., & Saß, H (2007). Mindestanforderungen für Schuld-fähigkeitsgutachten. [Minimum requirements for criminal responsibility assessments]. *Forensische Psychiatrie, Psychologie, Kriminologie, 1*, 3-9. https://doi.org/10.1007/s11757-006-0002-8

Feldhammer-Kahr, M., Kaiser, N., Leibetseder, I., Sommer, M., & Arendasy, M. E. (2024). Risk appraisal and legal principles - Unveiling disciplinary gaps. In C. Pracana, & M. Wang (Eds.), *Advances in Psychology and Psychological Trends, 2024* (pp. 305-315). Lissabon: inSciencePress. https://doi.org/10.36315/2024padX25

Diagnostic and Test Commission. (2017). *Qualitätsstandards für psychologische Gutachten* [Quality standards for psychological expert reports]. Berlin: Federation of German Psychologists' Associations.

Federal Ministry of the Interior, Amendment to the Ordinance of the Federal Minister of the Interior, which modifies the 1st and 2nd Weapons Act Implementation Ordinances. Retrieved from https://www.bmi.gv.at/401/files/2026/WaffV/Text_bf_20260204.pdf

Fegert, J. M., Gerke, J., Kliemann, A., Pusch, M., Rixen, S., & Sachser, C. (2024). *Die Methode der forensischen Glaubhaftigkeitsbegutachtung im deutschen Sprachraum. – Ein interdisziplinäres Plädoyer für eine kritische Bestandsaufnahme zur Anwendung der sogenannten 'Nullhypothese' in unterschiedlichen Verfahrenskontexten*. [The Method of forensic credibility assessment in German-speaking countries: An interdisciplinary call for a critical review of the application of the so-called 'Null Hypothesis' in Different Procedural Contexts.] Retrieved from: https://beauftragte-missbrauch.de/mediathek/publikationen/expertisen-und-studien

Hu, N., Pavlou, P. A., & Zhang, J. (2017). On Self-Selection Biases in Online Product Reviews. *MIS Quarterly*, *41*(2), 449-471. https://doi.org/10.25300/MISQ/2017/41.2.06

Kannegießer, A., Ebner, E., Wegmann, U., Grunert, S., Belke, A.-B., & Pfundmair, M. (2021). Peer-Review im Gutachterwesen - Wie kollegiales Feedback die Qualität familienpsychologischer Gutachten zu verbessern hilft. [Peer review in expert assessment: How collegial feedback helps improve the quality of family psychological expert reports.] *Psychologische Rundschau*, *72*(2), 147–149.

Kannegießer, A., Wegmann, U., & Ebner, E. (2020). *Abschlussbericht zum Pilotprojekt Professionelle Selbstkontrolle Online-Peer-Review-Verfahren* [Final report on the pilot project 'Professional self-monitoring: online peer-review procedure']. Münster: Kompetenzzentrum für Gutachten.

Kastner, P. (2019). Mindeststandards für forensisch-psychiatrische und psychologische Gutachten. [Minimum standards for forensic psychiatric and psychological assessments] In G. Brinek. (Ed.), *Gutachten als Schlüsselfaktoren im Maßnahmenvollzug. Schriftenreihe der Volksanwaltschaft* [Expert witness reports as key factors in the correctional system]. Series of publications from the Austrian Ombudsman]. Vienna: Austrian Ombudsman.

Kobbé, U. (2021). Forensische Prognosestellung nach dreijähriger Unterbringung im Maßregelvollzug – Herr L., 28 Jahre. [Forensic risk assessment after three years of detention in a psychiatric hospital – Mr. L., 28 years old] In K. D. Kubinger & T. M. Ortner (Eds.), *Psychologische Diagnostik in Fallbeispielen* [Psychological assessment in case studies] (pp. 400-415). Göttingen: Hogrefe.

Lichtenberger, E. O. (2006). Computer utilization and clinical judgement in psychological assessment reports. *Journal of Clinical Psychology*, *62*(1), 19-32. https://doi.org/10.1002/jclp.20197

Schömig, W. & Vogel, H. (2025). Systematische Qualitätssicherung und Qualitätsmanagement in der Begutachtung. [Systematic quality assurance and quality management in expert assessment]. In R. Dohrenbusch (Ed.) *Psychologische Begutachtung* [Psychological assessment]. (pp. 163-172). Berlin: Springer. https://doi.org/10.1007/978-3-662-64797-4_12

Tadei, A., Finnilä, K., Reite, A., Antfolk, A., & Santtila, P. (2016): Judges' Capacity to Evaluate Psychological and Psychiatric Expert Testimony. *Nordic Psychology*, *68*, 204-217. https://doi.org/10.1080/19012276.2015.1125303

Westhoff, K., & Kluck, M. L. (2014). Hilfen zur Beurteilung psychologischer Gutachten durch Fachfremde. [Guidelines for evaluating psychological expert reports by non-specialists]. In K. Westhoff, & M. L. Kluck (Eds.), *Psychologische Gutachten schreiben und beurteilen*. [Writing and evaluating psychological expert reports] (pp. 245-253). Berlin: Springer.