

Professional expectations about quality assurance: a review-based taxonomy of usability criteria in prevention, health promotion and education

Thomas Kliche · Jürgen Töppich · Stephan Kawski ·
Lutz Brunecker · Jale Önel · Anneke Ullrich ·
Uwe Koch

Received: 1 August 2005 / Accepted: 6 July 2006 / Published online: 5 October 2006
© Springer-Verlag 2006

Abstract Usability criteria sum up cognitions and perceptions of professional methods and problem solutions such as quality assurance in health care. They are important for the dissemination and implementation of innovations and the effectivity and efficiency of systems. In order to extract such criteria in an explorative structured review, publications from five fields were examined: (a) quality assurance in health care, (b) health promotion and education, (c) psychological methodology, (d) evaluation studies, and (e) comparative policy assessment. One hundred thirty-eight English and German basic and applied research contributions were selected. Usability criteria and comments on their interference were extracted. The review provided a two-dimensional taxonomy of the concept of usability. The first dimension is defined by five functional aspects of quality assurance instruments: (1) easy and efficient handling, (2) scope of practical functions and options for usage, (3) completeness of information, (4) users' compliance and motivation, and (5) scientific quality of results and data. Between these usability areas some characteristic antagonisms are perceived. The second dimension is defined by four levels of

increasing precision of professional cognitions: The usability area constitutes level 1. Each consists of several main criteria (level 2). Twenty-one main criteria of usability were found that were each specified by alternative sub-criteria (level 3) and their empirical indicators (level 4). This taxonomy offers an approach to the empirical exploration of users' expectations concerning quality assurance. It may contribute to the transparency of stakeholders' perspectives, to the preparation of consensus procedures in QA and to systematical comparison of quality assurance systems.

Keywords Quality assurance · Evaluation · Professions · Health promotion and education · Review

Usability criteria: professional cognitions framing health care quality assurance

The improvement of quality is a crucial issue in health care. Procedures of quality assurance (QA) have become a standard strategy to reach this goal, but they generate specific problems. Prevention, health promotion and education (HPE) can be considered as a case in point. There is considerable need for quality improvement in this field, because many programs, projects and interventions do not comply with professional standards. Process and outcome evaluations are frequently missing, evidence-based quality criteria are disregarded, process monitoring required for an adequate implementation is lacking in a lot of programs, and the introduction of QA routines is still pending, at least in Germany (e.g., Johnson et al. 2005; Kilian et al. 2004; Kirschner et al. 2004; Kliche et al. 2006; Marks and Sykes 2002; Molleman 2005). This critical description of HPE may be valid for many countries, because the International Union of Health Promotion and Education (IUHPE)

T. Kliche · S. Kawski · L. Brunecker · J. Önel · A. Ullrich ·
U. Koch
University Hospital Eppendorf (UKE), Medical Psychology,
Martinistr. 52 (S 35),
20246 Hamburg, Germany

J. Töppich
Federal Centre for Health Education (BzGA),
Ref 2–25, Ostmerheimer Str. 220,
51109 Cologne, Germany

T. Kliche (✉)
UKE, Medical Psychology,
Martinistr. 52 (S 35),
20246 Hamburg, Germany
e-mail: t.kliche@uke.uni-hamburg.de

initiated a project to integrate internationally available QA tools in order to put more evidence into practice (Van den Broucke 2005).

The transfer of adequate QA approaches from other fields of health care to HPE is a difficult task. QA approaches differ substantially with regard to the scope and strictness of criteria, the discretionary authority left to the practitioner, the professional target groups involved, the individual and organizational activities suggested, the resources and dissemination strategies. There are “internal” QA approaches (like self-evaluation), “professional” approaches (like guidelines and professional qualification), and “external” approaches, which can be distinguished into four main types: visitation, accreditation, EFQM and ISO assessments (Shaw 2003). For HPE, some authors advocate an “internal” QA approach (Wright 2004), while others already have developed systems for standardized external quality assessments (Kliche et al. 2004; Molleman et al. 2006).

The choice between these alternative approaches is difficult because the evidence on effects of QA procedures is minimal. Although instruments and designs for an evaluation of QA are available (Eccles et al. 2003; Hulscher et al. 2003; Tzavaras Catsambas et al. 2002; Wagner et al. 1999), there are few experimental studies on the comparative performance of QA approaches (McLoughlin and Leatherman 2003; McNamara 2005; Ovretveit and Gustafson 2002; Shaw 2003; Sheldon 2005). A review of 36 reviews yielded only seven randomized controlled trials of heterogeneous interventions (Wensing et al. 2006, 22 February). Three studies found no effects, while the other four indicated that only some specific components of quality management were effective (in a qualitative summary). The review concludes that “effects of quality management on professional performance and patient outcomes remain uncertain” (Wensing et al. 2006, 22 February, p. 12). Audit criteria of QA systems are only partly published and reveal a broad array of target dimensions, of which not all coherently focus on health care aspects (Hearnshaw et al. 2003). Less than 1% of the pertinent publications provides information both on the costs of implementing QA and on resultant changes in costs of care, which is a central comparative indicator and a justification for the QA-related additional work burden within an organization (Kilpatrick et al. 2005). Findings about the objectivity, reliability and validity of QA systems in health care have been rarely published; reasons for this gap may be methodological flaws of quality improvement research, low academic reward, reviewers’ unfamiliarity with improvement goals and methods, and publication guidelines (Baker 2006; Davidoff and Batalden 2005). However, another reason may be that there is no conclusive evidence supporting some rather expensive QA procedures.

Given these circumstances, the QA of QA is becoming increasingly important (AGREE-Collaboration 2003; Bird

et al. 2005; Helou et al. 2002). Currently in many health care institutions decisions regarding QA are based upon stakeholder representatives’ judgements. However, managers and professionals make different use of evidence and related resources for quality improvement (Lorenz et al. 2005). Consequently, they propose different sets of quality criteria (Campbell et al. 2004; Shield et al. 2003). However, expectations and role definitions of the professionals involved are a strong factor towards successful QA. It seems to be the interaction of professional self-regulation stimulated by internal QA measures and external QA methods that generate gains in quality, visibility and credibility for health care institutions by QA (Collopy 2000; Hunter et al. 2005; Johnston et al. 2000; Mittman 2004; Ovretveit 2003; Sheldon 2005; Van Weert 2000). If QA systems evoke misgivings concerning external control, unrealistic indicators, waste of time, and justifications for curbing budgets on behalf of the professional practitioners involved (Brezing 2000; Walshe et al. 2001), then the effectivity and efficiency of QA will be clearly reduced. Thus, professional expectations and cognitions are an important condition in the implementation of sustainable and successful QA in health care systems.

Professional protagonists’ cognitions guiding them to design, choose and implement innovations can be circumscribed as “usability criteria.” Sometimes they are referred to as “practicability,” but this is an ill-defined term that prolongs a misleading juxtaposition of academic thought and real-life orientation (Klein and Sorra 1996). Instead, the concept of usability criteria refers to the sociology and social psychology of knowledge. Professional expertise is a characteristic form of problem solving that integrates theory, methods and evidence, and that is guided by academic standards, practical experience, organizational culture and professional norms (Mieg 2001). If usability criteria can be conceptualized empirically, their contribution to innovations in health care can be appreciated. Usability criteria may be an emerging key issue of health care (Judd et al. 2001; Karsh 2004).

The objective of this study is therefore to address professional concepts and expectations that concern the design of adequate QA for HPE, because in this field there is no predominant QA system. Our method is to extract prevalent dimensions of professional judgements on QA from the relevant disciplines’ publications.

Methods

Structured review

Exploring heterogeneous fields of knowledge to collect relevant information can be done through a systematically structured Delphi-like approach, with a broad random

sample of sources and pragmatic consensus criteria (Dixon-Woods et al. 2004; Häder 2002). Thus, for a summary of experts' usability criteria, a review was designed comprising the relevant branches of knowledge and identifying patterns in their perspectives on QA. This research strategy follows theoretical sampling concepts in grounded theory (Strauss and Corbin 1996). The result is an overview of currently prevailing concepts.

The definition of relevant fields of knowledge

In order to represent the entire scope of professional perspectives, five main fields relevant to QA in HPE were chosen:

- (A) QA of health care in general.
- (B) QA for HPE.
- (C) Program evaluation and evaluation studies.
- (D) Psychological methodology (diagnostics and evaluation).
- (E) Political and administrative science and policy evaluation.

German and English publications were considered. These fields were selected because they represent the professions, functions and management, research and practitioners' perspectives cooperating in QA of HPE. (A) and (B) supply experience from different branches of health care; (C) to (E) complete this practical perspective by methodological knowledge from other disciplines. The combination of the included disciplines is based upon the assumption that the tasks, problems and methods of QA are similar throughout health care in an international perspective. Therefore, no distinction is made between prevention and health promotion.

Sources, inclusion and classification criteria

The five fields were investigated throughout the most important databases (Psycdex, PsycLit, Medline, PubMed, solis), the internet (Google, Metager), the Hamburg University Library and the institute's database (more than 900 titles covering mainly QA in rehabilitation). The thesauri only offered the possibility of circumscription, because usability criteria are discussed under different keywords and terms. As denominators, combinations of the following search dimensions had to be used (in German and English):

1. Field (examples of keywords: health promotion/education/prevention),
2. functions or tasks relevant to QA (quality/... assurance, ... development, ... measurement, ... assessment, etc.),
3. discussion of quality attributes (quality criteria/evaluation criteria/methodology/operationalization) or definitions of specific criteria (standardisation/validation/validity/objectivity/reliability).

Additionally, publications since 1970 retrieved from citations were included (snowball method). For the field of evaluation studies, the ten most frequently recommended publications (ten hits or more) recommended by about 200 experts in a Delphi study (Balzer 2003) were integrated into the review.

This search rendered over 200 heterogeneous publications. After exclusion of irrelevant hits, 138 German and Anglo-Saxon publications of basic and applied research were listed (cf. Table 1; list available from the authors). Among them there were many edited books, so that actually an even higher number of individual contributions and authors was included. The focus of most publications was the analysis of health care or politics; policy evaluation had environmental policy and administrative reform as additional central topics.

Two raters allotted the publications to the working fields and furthermore classified them either into basic contributions (postgraduate textbooks and key readings, introductory manuals and handbooks, methodological and basic readers) or applied research (reports and case studies on specific projects or problems). Basic contributions represent the current state of academic debate and are the fundament of professional knowledge about usability criteria, which have to be amended and modified in applied work in order to suit the individual task. The inclusion of these levels of discourse ensured that the central perspectives of professional knowledge were taken into account: scientific state-of-the-art and problem-oriented practical experience.

Some publications could be classified into different areas; in these cases, the more general one was chosen. An exact classification and a higher claim for completeness of publications retrieved and included was not necessary, because the goal of the study was a survey and not a consensus.

Analysis of the publications

As the publications thus varied systematically, it may be assumed that they cover the range of experts' views on QA. Publications from basic research imply a high degree

Table 1 Publications analyzed (functional area, number, kind of text, publication year)

Functional area	Basic contributions	Applied research
QA in health care	10 (1995–2002)	21 (1982–2004)
QA for HPE	11 (1995–2003)	7 (1992–2004)
Program evaluation und evaluation research	18 (1988–2003)	19 (1979–2003)
Psychologic diagnostics and evaluation	10 (1982–2002)	18 (1976–2003)
Political and administration science, policy evaluation	14 (1994–2004)	10 (1995–2004)

of theoretical and methodological consensus while publications from applied research represent usability considerations concerning specific professional tasks. The condensation technique of qualitative heuristics was used to identify common usability criteria for QA (Kleining 1995). Similar or doubtful criteria were discussed and then summarized and classified by two raters. In order to collect the entire scope of available knowledge, the conditions for extraction were defined liberally. For admission into the list of generally important criteria, a proposal had to be made

- in minimum five publications from
- in minimum two of the represented working fields.

If a criterion met these conditions, further quotations were not counted. A more detailed quantitative analysis as would be necessary for the ranking of evidence is informative only for specific empirical questions on the basis of a complete list of pertinent publications retrieved through broadly accepted denominators.

Additionally, hints from the texts concerning correlations, classification, interferences or conflicts of the criteria were marked. The reconstruction of these structural patterns resulted in a consistent taxonomy.

Results: a taxonomy of usability criteria

The extracted usability criteria establish a taxonomy consisting of two logically and operationally distinct dimensions. The first dimension is marked by five usability areas. They describe eminent functional aspects of QA systems. The second dimension results from the successive operationalization of usability on four levels of precision (cf. Table 2). The usability areas delineate the highest and most abstract level (level 1). They are delimited by main criteria that address central functional features of QA methods (level 2). On the next level, subcriteria answer specific questions or point out alternative options for how to comply with the main usability criteria (level 3). For example, the main criterion reliability in the area of the scientific or methodological quality of QA may refer to subcriteria like internal consistence, concordance, and retest-reliability.

These subcriteria can be operationalized and measured through several specific indicators or research methods (level 4).

The analysis of multiple quotations provided a delimited and manageable list of main criteria (cf. Table 3). Beneath the level of the main criteria (level 2), the concordance of the literature was marginal. Most of the subcriteria (level 3) were quoted in only one of the five publication areas. Even within academic disciplines, different definitions of subcriteria coexist [cf. “expert consensus,” “content validity” and “logic validity” of measurement instruments as a vague aim or as a basic scientific requirement (Bortz and Döring 2002; Jäger 1988; Michel and Conrad 1982)]. Some subcriteria are pertinent for several main criteria. Thus, the length of QA instruments is as important to enhance user-friendliness as it is for economic data gathering and analysis. For the level of operationalization, not much congruence was found, with the exception of psychometric indicators. Even within the disciplinary publication fields (cf. 0), most indicators obtained fewer than five nominations.

Many contributions point out characteristic antagonisms between the usability areas (e.g., Bengel and Koch 1988; e.g., Brezing 2000; Fowler 2001; Rossi et al. 1999):

- Scientific quality may demand complex instruments and methods that will collide with easy, efficient handling.
- For the same reason, simple handling can be hard to accord with completeness and reach of information, because completeness requires detailed documentation and data to check for confounders, selective bias, multilevel and unintentional effects of interventions.
- A multifunctional system for many usages can evoke users’ fears of economic and social control and diminish their acceptance and motivation to deliver valid data and implement the procedures and instruments.

On the other hand, some usability areas can be considered as complementary. The completeness and functional scope of a method harmonize. Easy handling heightens the motivation of users, and a high motivation can improve validity, reliability and objectivity of project documentation and data, i.e., scientific quality. Therefore, for the design and choice of QA methods, such potential collisions and

Table 2 Taxonomic levels of usability—example (main criterion “reliability”/subcriterion “consistence”)

Level 1: usability area	Level 2: main criterion	Level 3: subcriterion and definition by guiding question	Level 4: operationalization (indicators to answer the guiding question)
Scientific quality of QA results	Reliability of quality assessment	Consistence: Is the measurement of a quality dimension by several single items (e.g., ratings) unidirectional and homogeneous?	Cronbach’s Alpha, split-half

Table 3 Main and subcriteria in the five usability areas

Usability areas (dark) and main criteria	Subcriteria (examples)
(1) Easy and efficient handling Efficient implementation and use: data, implementation and handling of instruments and analyses are inexpensive and quick. User-friendliness, proximity to users' technical needs: methods and instruments are manageable, well structured, graphically well edited and easy to understand.	Resources required for data gathering and data acquisition Resources required for analysis Overall expenditure for establishing the system Overall impression of manageability Length of instruments Comprehensibility of instruments (instructions, questionnaires, manual, program, analyses and results), adequate and familiar technical terms Training and specific instruction are not necessary New software is not required
(2) Scope of practical functions and options for usage Feedback: the QA system provides results and recommendations that can be used for purposeful improvements of specific projects. Comparison among institutions/suppliers: the QA system provides quality indicators (benchmarks) to classify the project in the field's range of performance. Definition of working standards: the QA system establishes clear-cut basic professional criteria for good-practice projects. Sensibility and fairness towards particular practical contexts and requirements: the QA system harmonizes with the conditions of intervention or implementation in different institutions, settings, or projects. Scope of applications: the QA system can be used for different tasks and functions (e.g., for process evaluation, internal and external assessment, project classification, and monitoring). Integration and tuning of complex targets: the findings or indicators of the QA system balance divergent goals and objectives of projects and interventions.	Analyses deliver differentiated project description or profile Feedback of practical, specific recommendations Priority list of desirable improvements Availability of benchmarks Availability of norm data Fair comparisons between similar institutions, programs, projects or interventions Appropriate timing of updates (project documentations, quality assessments) Transparent good-practice criteria Distinction of adequate (standard) and improvable (professionally or scientifically deficient) working approaches Identification of problematic projects Criteria and method take account of complex practical conditions Suitability for different kinds of working contexts and institutions Quality criteria address genuine output and outcome (quality, health) and do not only reflect financing or size of the projects or programs considered Number of possible functions Experience of running the system for several functions Different data sets (e.g., self-evaluation, visitation, measurement of projects' structural attributes) Sensitivity to small-scale improvements and minor changes Consideration of competing aims or values in health care systems Consideration of providers' working frames (institutionals and legal regulations, financial sources) Consideration of realistic implementation strategies and limits for quality improvements
(3) Completeness of data and reach of information Causality of performance: the QA system identifies factors of success or failure of projects and interventions. Coverage of unintentional side effects Coverage of economic efficiency (return on investments) Coverage of all quality levels	Measures for monitoring project development QA comprises elements of detailed process descriptions linking causes and tracing back impact (e.g., compilation of case-studies) Broad description of effects Controlling for confounders, multilevel interferences and selection effects Systematic exploration of projects' and interventions' unintended consequences Comparative data on costs per unit (of intervention) Calculation of programs' short- and long-term health outcomes Data on structural quality Data on conceptual quality Data on procedural quality Data on output/outcomes

Table 3 (continued)

Usability areas (dark) and main criteria	Subcriteria (examples)
Sophisticated description and assessment of the quality dimensions	Detailed documentation Detailed scrutiny of specific intervention characteristics and quality criteria
Scope of working fields: the QA method yields appropriate results for several areas of application (settings, target groups, etc.).	Experience of running the QA system in several institutional fields Valid and tested for several health problems Valid and tested for several target groups Valid and tested for several settings Valid and tested for several types of interventions or projects Flexibility and adaptability of methods and instruments
(4) Compliance and motivation of users Professional reputation: authors, host or supplier of the QA system are well known and exert institutional, professional or scientific authority.	Tried and tested models or predecessors for the QA system Authors' reputation Authors' influence (positions) International circulation or implementation of the QA system Reputable organization implementing or coordinating the QA system
Transparent purpose: the QA system yields manifest gains of useful knowledge in foreseeable time.	Theoretical or conceptual reasonableness of the QA approach Practical reasonableness for the envisaged functions and usages Convincing design and elements
Manifest reference to the subject matter (face validity): it is evident that the QA system really addresses issues of quality.	Transparent aims of the QA system and its implementation QA instruments and methods are linked to the aims of QA QA outputs can be expected to foster quality improvements
Participation: all stakeholders are involved in the development and implementation of the QA system.	Transdisciplinary and multiprofessional cooperation Involvement of financing agencies and health care suppliers Representation of operators and users
(5) Scientific quality of results and data Validity: indicators and assessment results depict real qualities of projects and programs.	Content validity Predictive validity Ecological/external validity Population validity (e.g., for types of suppliers)
Reliability: QA results are precise, consistent and replicable.	Retest reliability Internal consistency
Objectivity: Data and analyses are generated free of interest collisions and improper influence.	Transparency of proceedings Impartiality of data gathering Impartiality of analysis

balances of usability areas have to be reflected and related to the aims and perspectives of stakeholders and users.

Discussion: limitations and options for using the taxonomy

The review succeeded in extracting a delimited number of distinct usability areas and criteria for QA in HPE. However, the following arguments have to be considered:

- (1) The taxonomy is both an inventory of features of QA systems and of professional and scientific dimensions to assess QA. It is not a normative checklist for minimal performance criteria of QA procedures. It may be used for the empirical description of experts' preferences, and the criteria can be adapted and weighted for specific purposes, for example, for the

construction of specific QA systems. In the selection of publications, it was assumed that applied research represents practical perspectives. However, for practical contexts, the relevance and interpretation of the usability criteria should be validated by users' views.

- (2) Usability areas and criteria differ in methodological weight. Only scientific quality ensures the significance of a method; easy handling, broad acceptance, etc., cannot compensate for failings in this area. Nevertheless, from the users' point of view, other priorities can be possible. A QA method with low validity can still be attractive to stakeholders, e.g., due to an expected participatory, educative and capacity-building impact. Therefore, QA systems are applied in HPE, although their scientific quality measures are in most cases completely unknown. Appraising the comparative significance of usability areas and criteria is a research task.

- (3) The taxonomy contains both generic criteria pertinent to measurement instruments as such (e.g., validity, economy, and comprehensibility) and characteristic criteria for functions of QA (feed-back and benchmarking). It is an open question which criteria have comprehensive and which have limited meaning. The taxonomy enables this question to be treated empirically. Through questioning of experts and users, specially tailored criteria lists and questionnaires for settings (schools, businesses, etc.) or for different branches of health care (prevention, rehabilitation, etc.) can be developed.
- (4) The publications' dissent about subcriteria and operationalizations shows that for the choice of indicators the specific aims and constraints of a given QA context has to be held in mind. Thus, according to some authors, the choice of criteria should depend on the project size (Harvey et al. 1992; Hulscher et al. 2003). For many contexts of HPE, the construction of valid and reliable indicators is still in the making.
- (5) Despite intensive and structured inquiry, the review still sums up only a sample of publications. Denominators and selected research areas showed some overlapping, and rehabilitation and psychology were more thoroughly represented than others. Further research might lead to other criteria. However, the structured approach was aimed at the inclusion of all relevant professional perspectives. The extracted criteria are very comprehensive on their own. In a validation study, 83 experts only sporadically nominated additional criteria. Thus, it is unlikely that more extensive research will lead to a broader spectrum of criteria. It would probably only generate more subcriteria and indicators.
- (6) It is an open question whether professional expectations concerning QA differ substantially among experts of several fields. If they do, a common taxonomy would be misleading, for example, in view of different approaches in preventive care and health promotion (Rosenbrock 2004). According to the review results, this is not very likely on the level of the main criteria, but distinct patterns might be hidden at the levels of subcriteria and operationalizations. If consensus procedures or other data indicate differing criteria in fields like primary care, prevention, health promotion, etc., several sets of usability criteria will have to be formulated. The taxonomy may be a start for closer scrutiny.
- The criteria may be preliminary and transitory. They cast a spotlight on current debates. As a consequence of professionalization in HPE and emerging standards for QA, increasingly refined and operationalized criteria can result.

Presently, individual decision-makers' preferences or small expert groups' deliberations control the choice, weighting and measurement of usability criteria for the specific QA tasks at stake. As an alternative, the taxonomy offers a framework for task orientation and transparency in the development of QA methodology. In the first place, the taxonomy provides a grid of criteria that helps to prepare systematic descriptions or comparisons of instruments and methods of QA. To try this, the taxonomy was used for the systematic description and comparative appraisal of several QA systems. Secondly, the taxonomy provides a device to bring into focus the implicit controversies in experts' and stakeholders' views on QA in general and on the propensities of given or planned QA systems in particular. To test this possibility, a questionnaire with the usability areas and main criteria was developed and validated by 83 experts of QA for HPE in Germany (a report and a study based on a sample of more than 200 experts is in preparation).

In summary, the taxonomy may contribute to the optimization of QA in health care by the following functions:

1. The questioning of experts about priorities, collisions and experience-based operationalizations of criteria can tap new resources of available professional knowledge.
2. The empiric inventory of expectations concerning QA makes professional positions transparent so that dissent can be handled more precisely (e.g., among health care suppliers, insurance companies and practitioners). Constraints and priorities for desirable QA designs can be defined precisely. The taxonomy may contribute to the validity and efficiency of consensus processes.
3. Differing ratings of experts about usability criteria mark the demand for further professional training and development of new QA instruments.
4. The homogeneity of peer reviewers' professional orientations in QA audits can be checked by analyzing their preference patterns for the criteria, as tested in a pilot study.
5. If the taxonomy is applied for comparisons of existing QA systems, profiles of strengths and weaknesses emerge and conditions of implementation for QA systems can be deduced.

Thus, the taxonomy enhances the quality of QA (Helou et al. 2002; Lehmann 2000). It can make expectations concerning QA and performances of QA systems transparent, in order to design innovations in a better way, implement them more effectively and efficiently and in the long

Because of these reasons, the taxonomy has two limitations of validity:

- It is based only on an exploratory review for QA of HPE. For other fields, it has to be reaffirmed by further evidence.

run transform the experience into professional standards for the QA of HPE.

References

- AGREE-Collaboration (2003) Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. *Qual Saf Health Care* 12(1):18–23
- Baker GR (2006) Strengthening the contribution of quality improvement research to evidence based health care. *Qual Saf Health Care* 15(3):150–151
- Balzer L (2003) Literaturliste zur Delphistudie: “Wie werden Evaluationsprojekte erfolgreich?” Universität Landau, Zentrum für empirische pädagogische Forschung, Landau
- Bengel J, Koch U (1988) Evaluationsforschung im Gesundheitswesen. In: Koch U, Lucius-Hoene G, Stegie R (eds) *Handbuch der Rehabilitationspsychologie*. Springer Berlin Heidelberg New York, pp 321–347
- Bird SM, Cox D, Farewell VT, Goldstein H, Holt T, Smith PC (2005) Performance indicators: good, bad, and ugly. *J R Stat Soc: Series A (Statistics in Society)* 168(1):1–27
- Bortz J, Döring N (2002) *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Springer Berlin Heidelberg New York
- Brezing H (2000) Welche Beduerfnisse haben Anwender(innen), und wie werden sie in der Forschung abgedeckt? Die Bedeutung von Evaluationsstandards und von Effektivitätskriterien für die Praxis. In: Hager W, Patry J-L, Brezing H (eds) *Evaluation psychologischer Interventionsmassnahmen. Standards und Kriterien: Ein Handbuch*. Huber, Bern, pp 8–18
- Campbell SM, Shield T, Rogers A, Gask L (2004) How do stakeholder groups vary in a Delphi technique about primary mental health care and what factors influence their ratings? *Qual Saf Health Care* 13(6):428–434
- Collopy BT (2000) Clinical indicators in accreditation: an effective stimulus to improve patient care. *Int J Qual Health Care* 12(3):211–216
- Davidoff F, Batalden P (2005) Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care* 14(5):319–325
- Dixon-Woods M, Agarwal S, Young B, Jones D, Sutton A (2004) Integrative approaches to qualitative and quantitative evidence. NHS, Health Development Agency, Public Health Evidence Steering Group
- Eccles M, Grimshaw J, Campbell M, Ramsay C (2003) Research designs for studies evaluating the effectiveness of change and improvement strategies. *Qual Saf Health Care* 12(1):47–52
- Fowler FJ (2001) Why it is easy to write bad questions. *ZUMA-Nachrichten* (48):49–66
- Häder M (2002) *Delphi-Befragungen. Ein Arbeitsbuch*. Westdeutscher Verlag, Wiesbaden
- Harvey RF, Silverstein B, Venzon MA, Kilgore KM, Fisher WP, Steiner M et al (1992) Applying psychometric criteria to functional assessment in medical rehabilitation: III. Construct validity and predicting level of care. *Arch Phys Med Rehabil* 73:887–892
- Hearnshaw HM, Harker RM, Cheater FM, Baker RH, Grimshaw GM (2003) Are audits wasting resources by measuring the wrong things? A survey of methods used to select audit review criteria. *Qual Saf Health Care* 12(1):24–28
- Helou A, Schwartz FW, Ollenschläger G (2002) Qualitätsmanagement und Qualitätssicherung in Deutschland. Übersicht auf der Grundlage des Gutachtens “Bedarfsgerechtigkeit und Wirtschaftlichkeit” des Sachverständigenrates für die Konzertierte Aktion im Gesundheitswesen 2000/2001. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, 45(3):205–214
- Hulscher MEJL, Laurant MGH, Grol RPTM (2003) Process evaluation on quality improvement interventions. *Qual Saf Health Care* 12(1):40–46
- Hunter WM, Schmidt ER, Zakocs R (2005) Evaluation of a peer assessment approach for enhancing the organizational capacity of state injury prevention programs. *J Public Health Manag Pract* 11(1):29–36
- Jäger RS (ed) (1988) *Psychologische Diagnostik: Ein Lehrbuch*. Weinheim: Psychologie Verlags Union
- Johnson HH, Glascoff MA, Lovelace K, Bibeau DL, Tyler ET (2005) Assessment of public health educator practice: health educator responsibilities. *Health Promot Pract* 6(1):89–96
- Johnston G, Crombie IK, Alder EM, Davies HTO, Millard A (2000) Reviewing audit: barriers and facilitating factors for effective clinical audit. *Qual Saf Health Care* 9(1):23–36
- Judd J, Frankish CJ, Moulton G (2001) Setting standards in the evaluation of community-based health promotion programs—a unifying approach. *Health Promot Int* 16(4):367–380
- Karsh B-T (2004) Beyond usability: designing effective technology implementation systems to promote patient safety. *Qual Saf Health Care* 13(5):388–394
- Kilian H, Geene R, Philippi T, Walter D (2004) Die Praxis der Gesundheitsförderung für sozial Benachteiligte im Setting. In: Rosenbrock R, Bellwinkel M, Schröer A (eds) *Primärprävention im Kontext sozialer Ungleichheit. Wissenschaftliche Gutachten zum BKK-Programm ‘Mehr Gesundheit für alle’*. Verlag für neue Wissenschaft, Bremerhaven
- Kilpatrick KE, Lohr KN, Leatherman S, Pink G, Buckel JM, Legarde C et al (2005) The insufficiency of evidence to establish the business case for quality. *Int J Qual Health Care* 17(4):347–355
- Kirschner R, Elkeles T, Kirschner W (2004) Evaluation des Fonds Gesundes Österreich. *Public Health Forum* 12(45):15
- Klein KJ, Sorra JS (1996) The challenge of innovation implementation. *Acad Manage Rev* 21(4):1055–1060
- Kleining G (1995) *Von der Hermeneutik zur qualitativen Heuristik. Lehrbuch entdeckende Sozialforschung, Bd. 1*. Weinheim: Beltz
- Kliche T, Töppich J, Kawski S, Koch U, Lehmann H (2004) Die Beurteilung der Struktur-, Konzept- und Prozessqualität von Prävention und Gesundheitsförderung: Anforderungen und Lösungen. *Bundesgesundheitsbl-Gesundheitsforsch-Gesundheitsschutz* 47:125–132
- Kliche T, Krüger C, Goldapp C, Mann R, Töppich J, Koch U (2006) Adipositas-Prävention für Kinder und Jugendliche in der Bundesrepublik - eine qualitätsorientierte Bestandsaufnahme. In: Kirch W, Badura B (eds) *Prävention. Ausgewählte Beiträge des Nationalen Präventionskongresses*. Springer, Berlin Heidelberg New York, pp 409–428
- Lehmann MA (2000) *Negotiating environmental quality: policy implementation in Germany and the United States*. Elgar, Cheltenham
- Lorenz KA, Ryan GW, Morton SC, Chan KS, Wang S, Shekelle PG (2005) A qualitative examination of primary care providers’ and physician managers’ uses and views of research evidence. *Int J Qual Health Care* 17(5):409–414
- Marks DF, Sykes CM (2002) Evaluation of the European Union Programme of Community Action on Health Promotion, Information, Education and Training 1996–2000. *Health Promot Int* 17(2):105–118
- McLoughlin V, Leatherman S (2003) Quality or financing: what drives design of the health care system? *Qual Saf Health Care* 12(2):136–142
- McNamara P (2005) Quality-based payment: six case examples. *Int J Qual Health Care* 17(4):357–362
- Michel L, Conrad W (1982) Theoretische Grundlagen psychometrischer Tests. In: Groffmann KJ, Michel L (eds) *Grundlagen psychologischer Diagnostik* pp 1–129. Hogrefe, Göttingen

- Mieg HA (2001) The social psychology of expertise. Case studies in research, professional domains, and expert roles. Erlbaum, Mahwah, NJ
- Mittman BS (2004) Creating the evidence base for quality improvement collaboratives. *Ann Intern Med* 140(11):897–901
- Molleman, GRM (2005) Preffi 2.0: Health Promotion Effect Management Instrument. Development, validity, reliability and usability. Woerden, NL: Netherlands Institute for Health Promotion and Disease Prevention (NIGZ)
- Molleman GRM, Peters LWH, Hosman CMH, Kok GJ, Oosterveld P (2006) Project quality rating by experts and practitioners: experience with Preffi 2.0 as a quality assessment instrument. *Health Educ Res* 21(2):219–229
- Ovretveit J (2003) What are the best strategies for ensuring quality in hospitals? WHO Regional Office for Europe's Health Evidence Network (HEN) synthesis report on the best strategies for ensuring quality in hospitals. WHO Regional Office for Europe, Copenhagen
- Ovretveit J, Gustafson D (2002) Evaluation of quality improvement programmes. *Qual Saf Health Care* 11(3):270–275
- Rosenbrock R (2004) Evidenzbasierung und Qualitätssicherung in der gesundheitsbezogenen Primärprävention. *Zeitschrift für Evaluation* 3(1):71–80
- Rossi PH, Freeman HE, Lipsey MW (1999) Evaluation: a systematic approach. SAGE Publications, Thousand Oaks; London; New Delhi
- Shaw CD (2003) Evaluating accreditation. *Int J Qual Health Care* 15(6):455–456
- Sheldon TA (2005) The healthcare quality measurement industry: time to slow the juggernaut? *Qual Saf Health Care* 14(1):3–4
- Shield T, Campbell S, Rogers A, Worrall A, Chew-Graham C, Gask L (2003) Quality indicators for primary care mental health services. *Qual Saf Health Care* 12(2):100–106
- Strauss A, Corbin J (1996) Grounded theory: Grundlagen qualitativer Sozialforschung. Beltz-Psychologie Verlags Union, Weinheim
- Tzavaras Catsambas T, Kelley ED, Legros S, Massoud R, Bouchet B (2002) The evaluation of quality assurance: developing and testing practical methods for managers. *Int J Qual Health Care* 14(90001):75–81
- Van den Broecke S (2005) Practice built on evidence. Guidelines and quality tools for European health. European Project Getting Evidence into Practice. Newsletter no 6. 6th IUHPE European Conference on the Effectiveness and Quality of Health Promotion, Best Practice for Better Health, Stockholm, p 4
- Van Weert C (2000) Developments in professional quality assurance towards quality improvement: some examples of peer review in the Netherlands and the United Kingdom. *Int J Qual Health Care* 12(3):239–242
- Wagner C, De Bakker DH, Groenewegen PP (1999) A measuring instrument for evaluation of quality systems. *Int J Qual Health Care* 11(2):119–130
- Walshe K, Wallace L, Freeman T, Latham L, Spurgeon P (2001) The external review of quality improvement in health care organizations: a qualitative study. *Int J Qual Health Care* 13(5):367–374
- Wensing M, Wollersheim H, Grol R (2006) Organizational interventions to implement improvements in patient care: a structured review of reviews. DOI 10.1186/1748-5908-1-2. *Implementation Science* 1(2):1–30, 22 February
- Wright M (2004) Partizipative Qualitätssicherung und Evaluation für Präventionsangebote in Settings. In R. Rosenbrock, M. Bellwinkel & A. Schröer (Eds), Primärprävention im Kontext sozialer Ungleichheit. Wissenschaftliche Gutachten zum BKK-Programm 'Mehr Gesundheit für alle' pp 297–346. Bremerhaven: Verlag für neue Wissenschaft