

Die Handlungsrelevanzmatrix als Baustein für die steuerungs- und entwicklungsorientierte Evaluation sozialer Dienstleistungen

Moisl, Dominique

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:

Moisl, D. (2011). Die Handlungsrelevanzmatrix als Baustein für die steuerungs- und entwicklungsorientierte Evaluation sozialer Dienstleistungen. *Zeitschrift für Evaluation (ZfEv)*, 10(1), 69-97. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-67412-2>

Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Die Handlungsrelevanzmatrix als Baustein für die steuerungs- und entwicklungsorientierte Evaluation sozialer Dienstleistungen

*Dominique Moisl*¹

Zusammenfassung: Auftraggeber/-innen stellen an Evaluationsforschung häufig die Anforderung, dass die Erkenntnisse direkt in Entscheidungs- und Steuerungsprozesse einfließen können. Die Handlungsrelevanzmatrix stellt eine leicht verständliche, grafische Auswertungsform für Befragungen dar, die es ermöglicht, unterschiedliche Aspekte des Leistungsgeschehens hinsichtlich des Leistungsniveaus (Performance) und hinsichtlich der Bedeutung (Relevanz) für den Erfolg der Dienstleistung zu analysieren. Mithilfe der Handlungsrelevanzmatrix können Praktiker/-innen auch ohne nennenswerte Methodenkenntnisse Handlungsbedarfe identifizieren und priorisieren. Die Methodik ist für die Evaluation und Verbesserung sozialer Dienstleistungen besonders gut geeignet, da sich mit ihr eine große Anzahl von Leistungsaspekten analysieren lässt. Darüber hinaus können Handlungsrelevanzmatrizen aufgrund der praxisorientierten Ergebnisdarstellung in Verbesserungsprozessen als Kommunikationsinstrument fungieren.

Schlagwörter: Handlungsrelevanzmatrix, Partizipative Evaluation, Praxisorientierte Ergebnisdarstellung, Handlungsempfehlungen

Importance-Performance-Analysis as a Module for Management-Oriented and Development-Oriented Evaluation of Social Services

Abstract: Clients often request that the findings of evaluation research can feed directly into decision-making and management processes. The Importance-Performance-Grid represents an easily understandable, graphic form of analysis of survey data, which makes it possible to analyze different aspects of service with regard to performance levels and importance (relevance) for the success of the service. The Importance-Performance-Grid helps practitioners – even without considerable knowledge of methods – to identify and prioritize needs for action. This methodology is very well suited for the evaluation and improvement of social services, because the analysis can cover a large number of aspects of service. In addition, the Importance-Performance-Grid can act as communication tool in improvement processes because of the practical presentation of results.

Keywords: Importance-Performance-Grid, Participatory Evaluation, Practical Presentation of Results, Recommendations for Action

1 Professur für „Sozialwissenschaftliche Aspekte der Sozialen Arbeit – Schwerpunkt Sozialplanung“ an der Fakultät für Allgemeinwissenschaften und Soziale Arbeit, Hochschule für angewandte Wissenschaften – Fachhochschule Landshut.

1. Anforderungen an steuerungs- und entwicklungsorientierte Evaluationsforschung

Auftraggeber/-innen von Evaluationsstudien haben verschiedene Erwartungen an den Nutzen einer Studie. Im Vordergrund steht häufig die Bewertung des Erfolges oder der Wirkungen einer Maßnahme (Rost 2000: 135; Menold 2007: 26ff.; Thayer/Fine 2001: 104ff.). Darüber hinaus sollen die Ergebnisse von Evaluationsstudien aber in vielen Fällen auch der Entwicklung und Verbesserung von Dienstleistungen oder Programmen dienen. Thayer und Fine (2001) führten in den USA eine Studie unter Non-Profit Organisationen aus dem Bereich der sozialen Dienstleistungen zu ihren Erfahrungen mit Evaluationsstudien durch. Als zentrale Nutzenaspekte einer Evaluation nannten die Befragten an erster Stelle die Bewertung und Dokumentation der Maßnahmenwirkungen, direkt gefolgt von der Identifikation von Ansatzpunkten für die Verbesserung der evaluierten Dienstleistungen. Evaluationsstudien waren den Befragten zufolge dann nützlich, wenn sie klare, spezifische Antworten und Empfehlungen für die Fragestellung der Auftraggebenden liefern konnten. Als besonders geeignet zur Deckung des spezifischen Informationsbedarfs der Auftraggebenden wurden Forschungsdesigns angesehen, welche Mitarbeiter/-innen aus der Praxis an der Studie bzw. der Formulierung der Fragebögen beteiligen und die Perspektiven der unterschiedlichen Stakeholder (relevanten Akteure) in die Evaluation einbeziehen (vgl. u.a. Thayer/Fine 2001: 106f.).²

Der vorliegende Beitrag stellt mit der Handlungsrelevanzmatrix einen methodischen Baustein für Evaluationen vor, der sowohl die hier von der Praxisseite genannten Aspekte für die Nützlichkeit von Evaluationen in der Konzeption, Auswertung und Ergebnisdarstellung berücksichtigt als auch weiteren zentralen Qualitätsanforderungen an Evaluationen in hohem Maße gerecht wird, wie z.B. den Standards der DeGEval – Gesellschaft für Evaluation e.V. (DeGEval 2008). Im Folgenden wird zunächst auf die Nützlichkeit als Qualitätskriterium von Evaluationen und die sich daraus ergebenden methodischen Anforderungen eingegangen. Anschließend wird das Konstruktionsprinzip der Handlungsrelevanzmatrix dargestellt und die Identifizierung und Priorisierung von Entwicklungsfeldern anhand eines Beispiels aus der Evaluationspraxis beschrieben. Der Beitrag schließt mit einer ausführlichen Diskussion methodischer Aspekte sowie der Eignung von Handlungsrelevanzmatrizen für die Verwendung in praktischen, steuerungs- und entwicklungsorientierten Arbeitskontexten.

1.1 Nützlichkeit als zentrales Kriterium der Evaluationsforschung

Die Nützlichkeit für den Auftraggeber stellt das zentrale Ziel von Evaluationen dar (vgl. Bewyl/Speer/Kehr 2004: 9) und ist gleichzeitig das zentrale Kriterium für die

2 Zu entsprechenden Ergebnissen kommt auch Stockbauer (2000a; 2000b), die anhand von drei Fallstudien bei Wiener Non-Profit-Einrichtungen und einer schriftlichen Befragung die Verwertbarkeit und Verwertung von Evaluationen untersucht.

Abgrenzung der Evaluationsforschung von der Grundlagen- bzw. disziplinentorientierten Forschung (vgl. Bortz/Döring 2005: 103ff.; Kromrey 2007: 115; Stockmann 2007a: 28f.). Im Rahmen von Evaluationsforschung sollen die Forschungsaktivitäten darauf ausgerichtet sein, dem Problemlösungsinteresse der Auftraggebenden gerecht zu werden und deren Fragen möglichst eindeutig und verständlich zu beantworten. Dem/der Evaluationsforscher/-in kommt dabei eine Ratgeberpflicht zu, d.h. er/sie muss eine klare Bewertung des Evaluationsgegenstands abgeben, damit die Evaluation für die Auftraggebenden in der jeweiligen Entscheidungssituation hilfreich ist (vgl. Bortz/Döring 2005: 103ff.). Evaluationen finden meist vor dem Hintergrund limitierter zeitlicher und finanzieller Ressourcen statt (vgl. u.a. Moisl 2010: 10). Dies bedeutet, dass Handlungsempfehlungen gegebenenfalls auch vor einem unvollständigen Informationshintergrund gegeben werden müssen. Für Evaluationsforscher/-innen ist es daher von besonderer Bedeutung, dass die Evaluationsergebnisse eine möglichst hohe Gültigkeit für die spezifische Fragestellung des jeweiligen Evaluationsprojektes haben (interne Validität) (vgl. Bortz/Döring 2005: 105).

Im Gegensatz dazu besteht der Auftrag der Grundlagen- bzw. disziplinentorientierten Forschung in der Generierung von allgemeinem Hintergrundwissen und unterliegt in der Regel keinem konkreten Verwertungsinteresse. Grundlagenforschende sind im Gegensatz zu Evaluierenden nicht der Erwartung ausgesetzt, bei Projektende eine abschließende Bewertung des Untersuchungsgegenstands abzugeben. Sie sind vielmehr angehalten, gemäß dem Postulat der „Wertfreiheit“ möglichst objektive Erkenntnisse zu generieren und die Studienergebnisse – unter Identifizierung bestehender Erkenntnislücken – selbstkritisch und vorsichtig zu interpretieren (vgl. Bortz/Döring 2005: 103ff.; Kromrey 2007: 113ff.). Dabei orientiert sich die grundlagenorientierte Vorgehensweise idealerweise (unter Ausblendung etwaiger Verwertungsinteressen) an rein wissenschaftsinternen Kriterien, mit dem Ziel, eine möglichst hohe Generalisierbarkeit (externe Validität) der Ergebnisse zu erreichen (vgl. Kromrey 2007: 114f.).

Auftrag und benötigte Informationen von Evaluations- und Grundlagenforschung unterscheiden sich somit deutlich. Zugespielt: die Evaluationsforschung benötigt möglichst situationsspezifische Erkenntnisse, die Grundlagenforschung strebt dagegen danach, von den Spezifika der konkreten Untersuchungssituation zu abstrahieren und Erkenntnisse mit situationsübergreifender Gültigkeit zu isolieren. Ein weiterer wichtiger Unterschied besteht in der Verwendung der Ergebnisse. Die Grundlagenforschung hat als Adressaten für ihre Forschungsergebnisse in erster Linie die Scientific Community und die Verbreitung der Ergebnisse geschieht hauptsächlich im Rahmen von wissenschaftlichen Publikationen (vgl. Kromrey 2007: 114). Die Evaluationsforschung generiert dagegen Wissen für Auftraggeber/-innen aus der Praxis, welches in vielen Fällen nicht öffentlich publiziert wird. Um den Wissenstransfer in die Praxis zu gewährleisten, müssen die Erkenntnisse so aufbereitet werden, dass sie auch für weitgehend evaluationsfremde Praktiker/-innen verständlich sowie im praktischen Arbeitskontext für Entscheidungsprozesse verwendbar sind bzw. dort auch tatsächlich verwendet werden. Dadurch besteht eine große Nähe zur Politikberatung und zur Organisationsentwicklung, insbesondere dann, wenn der Evaluationsauftrag die Ergebnisumsetzung bzw. die Initiierung

und Begleitung der Ergebnisumsetzung explizit einschließt (vgl. Kromrey 2007: 114; Müller 2009: 142ff.; Stockmann 2007b: 218)

1.2 Methodische Implikationen für die Evaluationsforschung

Die unterschiedlichen Rahmenbedingungen, Erkenntnisziele und Verwendungszwecke von Evaluations- und Grundlagenforschung berühren auch die Frage, welche Methoden angemessen sind, um die jeweiligen Untersuchungsziele zu erreichen. Kromrey (2007: 113) weist darauf hin, dass die Methoden der empirischen Sozialforschung und deren Qualitätskriterien für die Grundlagenforschung und die disziplinerorientierte Forschung entwickelt wurden. Die Evaluationsforschung bedient sich dieses Instrumentariums für die Informationsgewinnung allerdings vor dem Hintergrund der dargestellten, abweichenden Rahmenbedingungen und Erkenntnisziele. Evaluation gerät dabei immer wieder in Schwierigkeiten, Qualitätskriterien und Vorgehensweisen aus der Grundlagenforschung einzuhalten und auf dieser Basis eine für das spezifische Entscheidungsproblem des Auftraggebers nützliche, verständliche und verwertbare Informationsgrundlage zu schaffen (vgl. Stockmann 2007a: 29). Evaluation wird daher von Cronbach (1982) als „Kunst des Möglichen“ gesehen, bei der sich die methodische Ausgestaltung dem Ziel der Nützlichkeit für den Auftraggeber unterzuordnen habe. Mittlerweile wird jedoch weitgehend die Auffassung vertreten, dass sich Evaluationsforschung zwingend an den grundlegenden wissenschaftlichen und methodischen Standards der Grundlagenforschung orientieren müsse, um valide Ergebnisse zu erzielen (vgl. z.B. Bortz/Döring 2005: 102ff.; Stockmann 2007a: 28). Das Dilemma bringt Kromrey (2007: 115) wie folgt auf den Punkt: „(...) falsche – also empirisch nicht haltbare – Informationen sind genauso wertlos wie dies pragmatisch unbrauchbare Resultate wären.“

In der Evaluationspraxis schlagen sich die unterschiedlichen Anforderungen, insbesondere aus den Spezifika des Evaluationsauftrags, den allgemeinen Kriterien guter wissenschaftlicher Arbeit und den Qualitätsstandards für Evaluationen (z.B. DeGEval 2008) in einer Vielzahl sehr unterschiedlicher Evaluationsansätze und Evaluationsdesigns nieder. Häufig handelt es sich bei den einzelnen Evaluationsstudien um einen wenig systematischen und wenig abgestimmten Mix methodischer Einzelmaßnahmen, in welchem unterschiedliche Datenerhebungsmethoden bei unterschiedlichen Stakeholdergruppen zur Anwendung kommen. Die Kombination verschiedener Vorgehensweisen und Instrumente kann den Forschungsprozess dabei so komplex werden lassen, dass eine integrierte, nachvollziehbare und für den Auftraggeber handlungsleitende Auswertung der erhobenen Informationen die Evaluationsforscher/-innen vor große Schwierigkeiten stellt (vgl. Schuster 2007: 171f.; Koch/Fertsch-Röver 2009: 123).

2. Handlungsrelevanzmatrizen als Analysetool für steuerungs- und entwicklungsorientierte Fragestellungen

In der betriebswirtschaftlich geprägten Marketing-, Organisations- und Dienstleistungsforschung existiert für steuerungs- und entwicklungsorientierte Fragestellungen das Analyseinstrument der Handlungsrelevanzmatrix³ (vgl. Schmidt 2008: 304; Stauss 1999: 13f.). Handlungsrelevanzmatrizen sind eine spezielle grafische Aufbereitungsform für Untersuchungsergebnisse, die darauf abzielt, diejenigen Aspekte eines Evaluationsgegenstands zu identifizieren, für die ein vorrangiger Entwicklungs- bzw. Verbesserungsbedarf besteht (vgl. Bruhn 2008: 288ff.; Sampson/Showalter 1999: 2; Martilla/James 1977). In einer Handlungsrelevanzmatrix werden verschiedene Einzelaspekte des Untersuchungsgegenstandes (z.B. Qualitätsaspekte einer Dienstleistung) grafisch in direkten Bezug zu den beiden zentralen entscheidungsrelevanten Dimensionen, Leistungsniveau (gut – schlecht) und Relevanz für den Erfolg (wichtig – unwichtig), gesetzt. Die grafische Ergebnisdarstellung orientiert sich an der Verwendbarkeit in der Praxis: auch methodische Laien können die Ansatzpunkte für Steuerungs- und Entwicklungsbedarfe leicht aus einer Handlungsrelevanzmatrix ableiten und auf dieser Grundlage Handlungsoptionen selbst entwickeln oder Handlungsempfehlungen externer Evaluationen zumindest hinsichtlich deren empirischer Begründung nachvollziehen (vgl. z.B. Bauer 1996: 126). Die Erhebung der Datengrundlage zur Erstellung von Handlungsrelevanzmatrizen erfolgt über ein relativ einfaches Befragungs- und Analysekonzept. Zum einen folgen die Fragebögen einer standardisierten Form, die jedoch inhaltlich flexibel ist, um situations- und stakeholder-spezifische Evaluationsforschung zu ermöglichen (vgl. Hermann 2006). Zum anderen werden einfache, gut verständliche statistische Methoden so kombiniert, dass das gesamte Verfahren ausreichend robust ist, um auch in praktischen Arbeitszusammenhängen valide, belastbare Ergebnisse zu generieren.

Martilla und James haben bereits 1977 die Konstruktion und Anwendung von Handlungsrelevanzmatrizen für die Entwicklung von Marketingstrategien beschrieben. In der Folgezeit wurde das Tool im Detail weiterentwickelt und auf unterschiedliche Fragestellungen angewendet. Der Kern des Konzeptes ist jedoch weitgehend unverändert geblieben (vgl. Sampson/Showalter 1999: 1). In die sozialwissenschaftlich geprägte Evaluationsforschung haben Handlungsrelevanzmatrizen bisher kaum Eingang gefunden, obwohl sie gerade für soziale, personenbezogene Dienstleistungen mit ihrer hohen Komplexität ein ausgezeichnetes Analyseinstrument darstellen. Ein Grund für die seltene Verwendung könnte darin liegen, dass sich die methodische Anlage deutlich von den klassischen, in den Sozialwissenschaften etablierten Messkonzepten und den zugehörigen Gütekriterien abhebt.

3 Andere Begriffe für die Handlungsrelevanzmatrix sind Importance-Performance-Analyse (IPA), Importance-Performance-Grid, Zufriedenheitsportfolio, Zufriedenheitsprofil, u.a. (vgl. z.B. Schmidt 2008: 304).

3. Die Konstruktionselemente einer Handlungsrelevanzmatrix

3.1 Der Fragebogen

Handlungsrelevanzmatrizen bauen auf einem vergleichsweise einfachen Fragebogenkonzept auf. Für die Erstellung einer klassischen Handlungsrelevanzmatrix sind lediglich zwei Bausteine im Fragebogen obligatorisch: Erstens ein Messinstrument zur Erhebung des Erfolgskriteriums und zweitens ein Messinstrument zur Erhebung des Leistungsniveaus einzelner Aspekte des Dienstleistungsgeschehens (z.B. in Form der Beurteilung der Performance oder der Zufriedenheit).⁴

Das Erfolgskriterium wird in der Regel als subjektive Einschätzung der Befragten im Fragebogen erhoben.⁵ Die Messung kann entweder über eine einzelne Frage (z.B. eine Frage zur Gesamtzufriedenheit) oder über einen Index (z.B. einen Zufriedenheitsindex) geschehen. In der einschlägigen Literatur wird für eine valide Erfolgsmessung aus methodischen Gründen üblicherweise die Verwendung eines Index empfohlen (vgl. z.B. Homburg/Giering 1996; Churchill 1979; Töpfer 2008a: 351f.).⁶ Indizes basieren auf mehreren Einzelfragen, die – idealerweise – alle relevanten Dimensionen des Erfolgskriteriums erfassen. Die Antworten auf die einzelnen Fragen werden anschließend zu einem Indexwert zusammengerechnet, der als Erfolgsmaß dient (vgl. z.B. Meyer 2007a: 214ff.; Schnell/Hill/Esser 1999: 166ff.).⁷

Da das Instrument aus der Marketingforschung stammt, finden sich in der Literatur hauptsächlich Beispiele von Handlungsrelevanzmatrizen, bei welchen die Kundenzufriedenheit als Erfolgskriterium fungiert. Inhaltlich ist die Methode Handlungsrelevanzmatrix jedoch nicht festgelegt. Für das Evaluationspersonal besteht die – sowohl methodisch als auch theoretisch anspruchsvolle – Aufgabe darin, das für die jeweilige Untersuchungsfragestellung geeignete Erfolgs- oder Zielkriterium entweder aus bereits vorhandenen Messinstrumenten auszuwählen oder in angemessener

4 In der Regel wird der gesamte Fragebogen weitere Teile umfassen, insbesondere einen ‚klassischen‘ Befragungsteil (z.B. zur Abfrage von Fakten) sowie einen Teil mit Variablen zur Untergruppenbildung (z.B. Sozio-Demographie, Abteilungs-, Standortvariablen).

5 Ein zentrales Element der Evaluation von Dienstleistungsqualität ist typischerweise die Befragung der Stakeholder selbst (z.B. Kundenbefragung) (vgl. Sampson/Showalter 1999: 2; Matzler/Sauerwein/Heischmidt 2003: 115f.). Auf diese Weise erhält man das direkteste Feedback zur Dienstleistung ‚voice of the stakeholders‘ (Hermann 2006: 6). Soweit geeignete Indikatoren und Daten vorliegen, können als Erfolgsmaß grundsätzlich aber auch ‚objektive‘ befragungsexterne Daten (z.B. Prozessdaten) verwendet werden. Bezüglich einer ausführlichen Diskussion subjektiver und objektiver Erfolgsmaße sei auf Bachmann (2007) verwiesen.

6 Dagegen hat Reicheld (2003) mit dem Net-Promoter-Score (NPS) ein in der Praxis weit verbreitetes und akzeptiertes Erfolgsmaß für die Kundenbeziehung auf der Basis eines einzelnen Items, der Frage nach der Weiterempfehlungsbereitschaft, entwickelt. Den zentralen Vorteil der Verwendung einer Frage sieht er in der Verständlichkeit für Praktiker/-innen. Eine sorgfältig ausgewählte Frage sei deutlich besser als komplexe, mehrdimensionale Herangehensweisen.

7 Indizes sind grundsätzlich von Skalen zu unterscheiden, auch wenn die Begriffe ‚Index‘ und ‚Skala‘ in den Sozialwissenschaften uneinheitlich verwendet werden (vgl. Greving 2007: 66; Schnell/Hill/Esser 1999: 166). Eine ausführliche Darstellung der Unterschiede findet sich in Fußnote 22.

sener Weise selbst zu operationalisieren (für soziale Dienstleistungen könnten dies z.B. sein: Klienten-, Angehörigen- oder Mitarbeiterzufriedenheit, Qualität der Zusammenarbeit, Wirksamkeit, Partizipation im Beratungsprozess, Reputation des Angebotes oder des Trägers etc.).⁸ So orientiert sich Tornow (2006; vgl. auch e/l/s 2004) im Rahmen einer Studie zur Effektivität von stationärer Erziehungshilfe für die Konstruktion des Erfolgskriteriums der Handlungsrelevanzmatrix z.B. an den allgemeinen Kategorien für Dienstleistungsqualität des SERVQUAL-Ansatzes (Parasuraman/Zeithaml/Berry 1988), der mittlerweile im Rahmen des ARCH-SECRET-Modells (Vaughan/Shiu 2001) um non-profit-spezifische Dimensionen erweitert wurde.

Das Leistungsniveau einzelner Aspekte der Dienstleistung wird anhand einer einfachen Fragenbatterie erhoben. Zu jedem Leistungs- bzw. Qualitätsaspekt wird eine Frage formuliert (Ein-Item-Messung),⁹ die die Befragten jeweils anhand derselben Beurteilungs- oder Zufriedenheitsskala bewerten.¹⁰ Diese vergleichsweise schlichte Vorgehensweise hat einen zentralen Vorteil: Es können sehr viele Einzel- und Detailspekte abgefragt und von den Befragten beurteilt werden (vgl. Drollet/Morrison 2001: 201). Dadurch ergibt sich in der Auswertung ein sehr breites, klares und zugleich detailliertes Bild der Leistung und der Ansatzpunkte für Verbesserungen (vgl. Bruhn 2008: 138f., 287f.). Mit detaillierten Ergebnissen lassen sich Entwicklungsmaßnahmen zielführend lenken, so dass die Gefahr von Fehlinterpretationen, fehlgeleiteten Verbesserungsmaßnahmen sowie Fehlinvestitionen verringert wird.

3.2 Elemente der Ergebnisdarstellung in der Handlungsrelevanzmatrix

Die erhobenen Daten zum Erfolgskriterium und zur Beurteilung der einzelnen Leistungsaspekte bilden die Grundlage der Handlungsrelevanzmatrix. Jede Dimension bestimmt eine Achse des (zweidimensionalen) Koordinatensystems¹¹ der Handlungsrelevanzmatrix. Im aufgespannten Merkmalsraum werden die einzelnen Leistungsaspekte entsprechend ihrer erfragten bzw. errechneten Messwerte in Bezug auf die beiden Dimensionen positioniert. Aus der Position eines Leistungsaspektes im

8 Ein Set weniger, adäquater Messinstrumente zentraler und allgemein formulierter Erfolgskriterien dürfte für die standardmäßige Evaluation (sozialer) Dienstleistungsprozesse in der Regel ausreichend sein. So wird z.B. in dem auf Handlungsrelevanzmatrizen aufbauenden Befragungskonzept (bzw. Stakeholder Management Konzept) von TNS Infratest mit „nur“ sechs zentralen Erfolgsindizes gearbeitet: Kundenbindung, Mitarbeiterbindung, Interne Zusammenarbeit, Management Evaluation, Public Service Quality, Unternehmensreputation (vgl. Sondervan 2008; Hermann 2008; Huber/Scharioth 2003: 7).

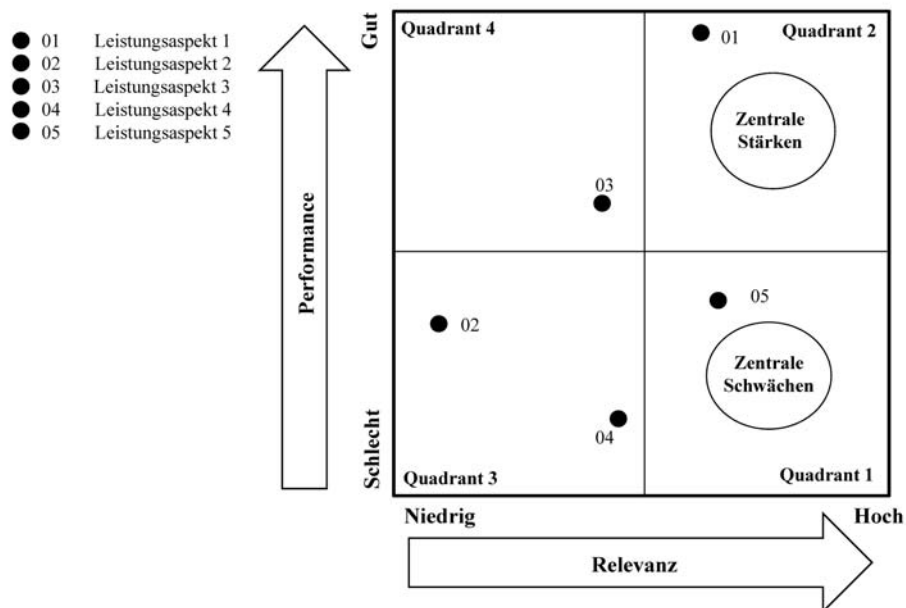
9 Eine ausführliche Diskussion von Ein-Item- und Mehr-Item-Messung findet sich in Kapitel 5.1.2.

10 Die entsprechende Frageformulierung könnte z.B. wie folgt aussehen: „Wie beurteilen Sie ... [Leistungsaspekt, z.B. die Freundlichkeit des Personals]?“ Eine mögliche Antwortskala wäre z.B. „sehr gut – gut – eher gut – eher schlecht – schlecht“.

11 Komplexere Handlungsrelevanzmatrizen arbeiten mit drei Dimensionen (z.B. Töpfer 2008a: 374; Sondervan 2008: 44f.). Neben der x- und der y-Achse des Koordinatensystems wird die dritte Dimension z.B. mittels unterschiedlicher Symbole dargestellt, die – anstelle einfacher Punkte – in der Matrix platziert werden (vgl. Sondervan 2008: 44f.).

Koordinatensystem lässt sich dann direkt auf dessen Verbesserungsbedarf schließen. Zur Vereinfachung der Interpretation wird der Merkmalsraum meist in Felder aufgeteilt oder mit anderen Markierungen versehen, die als Referenzpunkte dienen. Abbildung 1 zeigt das Konstruktionsprinzip einer Handlungsrelevanzmatrix, das im Folgenden anhand der verschiedenen Konstruktionselemente näher erläutert wird.

Abbildung 1: Schema einer Handlungsrelevanzmatrix



3.2.1 Koordinatenachsen

Die beiden Koordinatenachsen repräsentieren die zwei Entscheidungsdimensionen, nach welchen die einzelnen Leistungsaspekte auf ihre Handlungsrelevanz hin untersucht werden. Dies sind klassischerweise die beiden Dimensionen Performance (Leistungsniveau) und Relevanz (Wichtigkeit) (vgl. Martilla/James 1977):

1. *Performance (Leistungsniveau)*: zeigt die Beurteilung des Ist-Zustandes („Was läuft gut, was läuft schlecht?“ bzw. „Was wird – von unseren Klienten und Klientinnen, Kunden und Kundinnen, Mitarbeitern und Mitarbeiterinnen etc. – gut oder schlecht bewertet?“).
2. *Relevanz (Wichtigkeit)*: zeigt die Wichtigkeit der einzelnen Leistungsaspekte für das Erfolgskriterium bzw. die Zielerreichung („Was ist wichtig/unwichtig für den Erfolg/die Zielerreichung?“).

Die Belegung der Achsen des Koordinatensystems wird unterschiedlich gehandhabt. In der angelsächsischen Literatur wird die Relevanz häufig auf der vertikalen Achse, die Performance auf der Horizontalen abgetragen (vgl. z.B. Matzler/Sauerwein/Heischmidt 2003: 115; Martilla/James 1977: 78; aber auch Homburg/Werner

1998: 99). Im deutschsprachigen Raum findet oftmals die umgekehrte Achsenbelegung Verwendung (vgl. z.B. Töpfer 1996: 264; Becker 2009; Bruhn 2008: 289).

3.2.2 Positionierung der Leistungsaspekte

Die Positionierung im Koordinatensystem stellt alle Leistungselemente in einen relativen Bezug zueinander. Aus dieser Darstellung ergibt sich die Handlungspriorisierung (vgl. Homburg/Werner 1998: 91). Die zugrunde liegende Logik beschreibt Slack (1994: 62) wie folgt: „The assumed form of the relationship is that importance and performance act together to determine «priorities of improvement».“

Es existieren verschiedene Vorgehensweisen zur Messung der empirischen Werte für die beiden Dimensionen Leistungsniveau und Relevanz, die zur Positionierung der Leistungsaspekte im Koordinatensystem herangezogen werden.

Die Messwerte des Leistungsniveaus der einzelnen Leistungsaspekte werden fast immer direkt bei den relevanten Stakeholdern (z.B. den Kunden/Kundinnen, Klienten/Klientinnen) mithilfe eines Fragebogens erhoben. Diese direkte Messung hat gegenüber indirekten Ansätzen über Indikatoren oder Fremdurteile (z.B. die Einschätzung der Kundenzufriedenheit durch Mitarbeiter/-innen) den erheblichen Vorteil, dass sie weniger anfällig für Messfehler und Fehldeutungen ist. Bei indirekten Messungen besteht eine größere Gefahr von Fehleinschätzungen, die zu kostspieligen Fehlentscheidungen führen können (vgl. Töpfer/Mann 1996: 49). Für die Positionierung der Performancemessung wird häufig der Mittelwert der Beurteilungen eines Leistungsaspektes herangezogen (vgl. z.B. Bruhn 2008: 288; Sampson/Showalter 1999: 2ff.; Martilla/James 1977: 79). Zur Verwendung kommen aber auch der Median (vgl. Martilla/James 1977; Sampson/Showalter 1999: 3) oder Häufigkeitsanteile in bestimmten Antwortkategorien (z.B. die summierten Häufigkeiten in den beiden besten Antwortkategorien, wie etwa „vollkommen zufrieden“ und „sehr zufrieden“ bei Tornow 2006: 9ff.).

Zur Bestimmung der Relevanz der Leistungsaspekte für den Erfolg bzw. die Zielerreichung können verschiedene Wege eingeschlagen werden:

- a) *Expertenurteil*: Für die Bestimmung der Wichtigkeiten einzelner Leistungsaspekte für das Erfolgskriterium werden Experten/Expertinnen befragt (z.B. Mitarbeiter/-innen schätzen die Wichtigkeit einzelner Aspekte aus Klientensicht ein) bzw. wird die Literatur zu ähnlichen Studien gesichtet. Diese Vorgehensweise ist insbesondere dann von Bedeutung, wenn keine andere Möglichkeit zur Bestimmung der Relevanz gegeben ist (z.B. wenn im Nachhinein zu einer Befragung noch eine Handlungsrelevanzmatrix erstellt werden soll, Daten zur Relevanz aber fehlen). Expertenurteile haben jedoch neben der Gefahr von Fehleinschätzungen den Nachteil, dass auch unter Expertinnen und Experten in der Regel kein umfassendes Wissen über die Wichtigkeit einzelner Leistungsaspekte im konkret untersuchten Dienstleistungssetting vorhanden ist. Expertenurteile, ebenso wie Ergebnisse aus der Literatur, können deshalb in aller Regel nur allgemeines Wissen zu durchschnittlichen Wichtigkeiten bzw. zu Wichtigkeiten in ähnlichen Kontexten geben.

- b) *Direkte Abfrage der Relevanz im Fragebogen* (analog zur Vorgehensweise bei der Messung des Leistungsniveaus) (vgl. z.B. Sampson/Showalter 1999: 3):¹² Die direkte Abfrage von Wichtigkeiten im Fragebogen ist umstritten, da sich die verbalisierten Wichtigkeiten teilweise deutlich von statistisch bestimmten Wichtigkeiten unterscheiden (vgl. Jarvis 2008: 6; Matzler/Sauerwein/Stark 2008: 334; Stauss 1999: 13). Die direkte Abfrage von Wichtigkeiten ist mit verschiedenen methodischen Problemen verknüpft: So werden beispielsweise sozial erwünschte Attribute (z.B. Umweltfreundlichkeit) von Befragten in der Bedeutung für das eigene Verhalten häufig überschätzt, wohingegen selbstverständliche Leistungsattribute bzw. Basisfaktoren der Leistung (z.B. Einhaltung gesetzlicher Bestimmungen) häufig unterschätzt werden (vgl. Homburg/Werner 1998: 89). Ferner besteht eine grundsätzliche Tendenz, alle Faktoren im oberen Bereich der Wichtigkeit einzuordnen (Anspruchsinflation) (vgl. Dichtl/Müller 1986: 233). Ein zentraler praktisch-methodischer Nachteil der direkten Wichtigkeitsabfrage ist jedoch, dass zu jedem Leistungsattribut neben dem Leistungsniveau auch die Wichtigkeit erhoben werden muss. Dies verlängert den Fragebogen deutlich. Die direkte Abfrage von Wichtigkeiten kommt mittlerweile meist nur noch dann zur Anwendung, wenn die subjektive Sichtweise der Befragten eine eigenständige Untersuchungsfragestellung darstellt: „Stated importance isn't wrong, it's just a different form of reality“ (Jarvis 2008: 6).¹³
- c) *Statistische Berechnung der Relevanz*: Mittlerweile ist die statistische Berechnung der Wichtigkeit der einzelnen Leistungsattribute für das Erfolgskriterium die am besten akzeptierte und eine weit verbreitete Vorgehensweise (vgl. z.B. Homburg/Werner 1998: 90; Bruhn 2008: 289). Häufig wird dabei auf einfache Korrelationskoeffizienten zurückgegriffen (vgl. z.B. Jarvis 2008: 5f.), oft auf *Pearsons r*. *Pearsons Produkt-Moment-Korrelationskoeffizient r* findet auch bei ordinalem Skalenniveau Verwendung, setzt korrekterweise aber metrisches Skalenniveau voraus. Bei einigen Studien kommt daher der methodisch angemessenere, aber weit weniger gängige, *Spearman's Rangkorrelationskoeffizient Rho* zur Anwendung (vgl. Sampson/Showalter 1999: 4, 12). Wenige Studien verwenden auch Regressionsanalysen, Conjoint-Analysen oder kausal-analytische Vorgehensweisen zur Bestimmung der Relevanz (vgl. Stauss 1999: 14; Sampson/Showalter 1999; Homburg/Werner 1998: 90).¹⁴

12 Die entsprechende Frageformulierung könnte z.B. bei einer Befragung zur Klientenzufriedenheit wie folgt aussehen: „Wie wichtig ist für Sie ... [Leistungsaspekte, z.B. Freundlichkeit Ihres Betreuers]?“ Eine mögliche Antwortskala wäre z.B. „sehr wichtig – wichtig – eher wichtig – eher unwichtig – unwichtig“.

13 Eine interessante Analyseform bildet in diesem Zusammenhang das sogenannte Importance-Grid, in welchem – analog zur hier dargestellten Handlungsrelevanzmatrix – die verbale Wichtigkeit mit der errechneten Wichtigkeit in Beziehung gesetzt wird (vgl. Matzler/Sauerwein/Stark 2008: 336ff.; Matzler/Sauerwein 2002: 319ff.).

14 Für die Diskussion der Verwendung von Korrelationen zur Relevanzmessung vgl. Kapitel 5.2.5.

3.2.3 Referenzmarkierungen

Um die Interpretation der Positionierung zu erleichtern und zu standardisieren, werden in das Koordinatensystem Markierungen eingezogen. Klassischerweise wird die Matrix in 4 Felder eingeteilt, d.h. jede Achse erhält eine Markierung, die diese in einen Bereich mit ‚hohen‘ und mit ‚niedrigen‘ Werten einteilt (2 Felder je Achse: 2x2-Felder-Matrix; vgl. z.B. Martilla/James 1977; Jarvis 2008: 5; Becker 2009; Tornow 2006: 10f.). Für jedes Feld ist eine sogenannte „Normstrategie“ definiert, d.h. die Felder sind mit einer leitlinienartigen Interpretation hinsichtlich ihrer Handlungsrelevanz vorbelegt, so dass sich aus der Positionierung eines Leistungselements die Handlungsimplication ergibt (vgl. z.B. Sampson/Showalter 1999: 2).¹⁵

Zur Festlegung des Koordinatenkreuzes werden üblicherweise die arithmetischen Mittelwerte der Scores auf der jeweiligen Dimension verwendet. Andere Möglichkeiten sind zum Beispiel der Median (vgl. Sampson/Showalter 1999: 2f.; Tornow 2006: 10f.) oder normative Werte, wie z.B. ein angestrebtes Leistungsniveau oder das Leistungsniveau des Wettbewerbs (vgl. Slack 1994: 60). Die Festlegung der Grenzen der Quadranten ist somit etwas willkürlich und abhängig von der forschenden Person (vgl. Sampson/Showalter 1999: 3; Schmidt 2008: 304; Matzler/Sauerwein/Stark 2008: 339). Bei der Interpretation der Positionierung eines Leistungselements – Zugehörigkeit oder Nicht-Zugehörigkeit zu einem Quadranten – ist daher zu berücksichtigen, dass insbesondere die errechneten Grenzen lediglich als Orientierungsmarkierungen anzusehen sind und keine trennscharfe Klassifizierung darstellen. Die eigentliche Interpretation ist relativ in Bezug auf die Position der einzelnen Elemente zueinander vorzunehmen (vgl. Sampson/Showalter 1999: 3).¹⁶

3.2.4 Normstrategien

In Bezug auf die aus der Positionierung in der Matrix abzuleitenden Handlungsimplicationen, d.h. die Normstrategien, besteht in der Literatur grundlegende Einigkeit. Unterschiede treten lediglich in Bezug auf die Formulierung der Normstrategien im Detail auf.¹⁷ Im Folgenden werden die Normstrategien, die sich aus einer Handlungsrelevanzmatrix ergeben, bei einer klassischen Einteilung in vier Quadranten beschrieben (siehe Abbildung 1). Die Reihenfolge der Quadranten spiegelt

15 Die grundlegende 2x2-Felderstruktur der Handlungsrelevanzmatrix wird von verschiedenen Autoren „verfeinert“. So verwenden z.B. Töpfer und Mann (1996: 53) eine 4x4-Felder-Matrix und Slack (1994: 69) sogar eine 9x9-Felder-Matrix. Bei diesen Matrizen ist die Interpretation jedoch deutlich komplizierter, da nicht für jedes Feld eine Normstrategie formuliert wird und die Normstrategien sich nicht an den Feldergrenzen orientieren. Vielmehr werden felderübergreifende Zonen definiert, für die jeweils eine Normstrategie angegeben wird (vgl. Slack 1994: 5; Töpfer 2008a: 345).

16 Die relative Interpretation der Positionierung entspricht dabei auch dem Qualitätsmanagement-Ansatz insofern, als dass unabhängig vom absoluten Qualitätsniveau immer (relative) Stärken und Schwächen bzw. Verbesserungsmöglichkeiten gegeben sind (vgl. Jarvis 2008: 6).

17 Bezüglich unterschiedlicher Formulierungen für die Normstrategien vgl. z.B. Töpfer 1996: 53; Töpfer/Mann 1996: 264; Schmidt 2008: 305; Homburg/Werner 1998: 187; Sampson/Showalter 1999: 3; Matzler/Sauerwein/Heischmidt 2003: 115; Slack 1994: 61; Bauer 1996: 131.

eine abnehmende Handlungsrelevanz für Verbesserungen wider (vgl. Martilla/James 1977: 78):

- *Quadrant 1* – Leistungsaspekte mit *hoher Wichtigkeit* und *niedriger Performance-Beurteilung*: Diese Leistungselemente zeigen zentrale Schwächen auf. Hier besteht sofortiger und dringlicher Handlungsbedarf: „Sofortmaßnahmen“ (Töpfer 2008b: 429; 1996: 264); „Concentrate Here“ (Martilla/James 1977: 78).
- *Quadrant 2* – Leistungsaspekte mit *hoher Wichtigkeit* und *hoher Performance-Beurteilung*: Die Leistungselemente in diesem Quadranten stellen zentrale Stärken dar. Aus strategischer Sicht ist bei diesen Elementen darauf zu achten, dass das aktuell hohe Leistungsniveau weiterhin aufrecht erhalten wird: „Bestandssicherungsmaßnahmen“ (Töpfer 1996: 264); „Differenzierungsmaßnahmen“ (Töpfer 2008b: 429); „Keep Up The Good Work“ (Martilla/James 1977: 78).
- *Quadrant 3* – Leistungsaspekte mit *niedriger Wichtigkeit* und *niedriger Performance-Beurteilung*: Bei den Leistungselementen in diesem Quadranten ist fallweise zu prüfen, ob – mit niedriger Priorität – Verbesserungen durchgeführt werden sollen. Verbesserungen können insbesondere dann erwogen werden, wenn diese einfach und mit geringem Aufwand durchführbar sind: Maßnahmen der „Basisabsicherung“ (Töpfer 2008b: 429; 1996: 264); „Low Priority“ (Martilla/James 1977: 78).
- *Quadrant 4* – Leistungsaspekte mit *niedriger Wichtigkeit* und *hoher Performance-Beurteilung*: Für Leistungselemente in diesem Quadranten besteht kein dringlicher Handlungsbedarf. Es ist zu prüfen, ob bei den hier positionierten Elementen Einsparpotential besteht: Maßnahmen für „Einsparpotential“ (Töpfer 2008b: 429; 1996: 264); „Possible Overkill“ (Martilla/James 1977: 78).

4. Beispiel: Auszug aus einer Handlungsrelevanzmatrix

Der folgende Auszug aus einer Handlungsrelevanzmatrix zeigt die Ergebnisdarstellung an einem konkreten Beispiel aus der Evaluationspraxis und verdeutlicht die Ableitung von Handlungsempfehlungen. Für die Erstellung des Beispiels wurde auf reale Daten aus einem Projekt zur Evaluation sozialer, personenbezogener Dienstleistungen – im Speziellen der Ko-Produktion im Krankenhaus – zurückgegriffen.

4.1 Datengrundlage

Datengrundlage ist eine Unterstichprobe der Kölner Patientenbefragung aus dem Jahr 2000. Insgesamt wurden 700 Patientinnen und Patienten der chirurgischen und inneren Fachabteilungen der Universitätsklinik Köln per Zufallsverfahren ausgewählt und postalisch mit Fragebögen in drei unterschiedlichen Längen zu ihrem Krankenhausaufenthalt befragt. Die Unterstichprobe umfasst 248 Patientinnen und

Patienten, die die längste, 12-seitige Version, des Kölner Patientenfragebogens ausgefüllt haben. Die Rücklaufquote lag bei 65 Prozent.¹⁸

Die gesamte Handlungsrelevanzmatrix beruht auf der Analyse von 54 Fragen zu Leistungsaspekten. Es handelt sich um 27 unterschiedliche Leistungselemente, die jeweils für Ärzte/Ärztinnen und Pflegekräfte erhoben wurden.¹⁹ Der dargestellte Auszug zeigt fünf Leistungselemente – jeweils für Pflegekräfte und Ärzte/Ärztinnen – die zentral für die Ko-Produktion sozialer Dienstleistungen sind (vgl. Abb. 2). Für die Bestimmung der Wichtigkeit der Leistungsaspekte wird der Zusammenhang mit der Gesamtzufriedenheit mit dem Krankenhausaufenthalt, welche als Erfolgskriterium fungiert, herangezogen.²⁰

Tabelle 1 zeigt die Messwerte zur Beurteilung der Performance und zur Wichtigkeit für die 10 ausgewählten Leistungsaspekte. Für alle 54 Aspekte, die die Datenbasis der Handlungsrelevanzmatrix bilden, liegt die Spannweite der mittleren Beurteilung (arithmetisches Mittel) zwischen 2,54 (Pflegekräfte führten regelmäßig Gespräche mit mir) und 3,64 (gerechte Behandlung durch Pflegekräfte) bei einem arithmetischen Mittel der Mittelwerte von 3,15. Hinsichtlich der gemessenen Wichtigkeit liegt die Spannweite der Korrelationen mit dem Erfolgskriterium zwischen $r=.13$ (Pflegekräfte haben zu oft gewechselt) und $r=.63$ (Ärzte/Ärztinnen waren offen und ehrlich zu mir) bei einem arithmetischen Mittel der Korrelationskoeffizienten von .44.

Tabelle 1: Messwerte, die der Handlungsrelevanzmatrix zugrunde liegen

	Pflegekräfte		Ärzte/Ärztinnen	
	Beurteilung Performance ^b	Relevanz Korr (r) ^a	Beurteilung Performance ^b	Relevanz Korr (r) ^a
Gerechte Behandlung durch P/Ä	3,643	.314	3,625	.473
P/Ä haben zu häufig gewechselt	2,774	.125	2,903	.149
Fachkompetenz der P/Ä	3,250	.386	3,468	.500
Erreichbarkeit der P/Ä	3,134	.388	2,691	.409
P/Ä führten Gespräche gefühlvoll durch	2,736	.408	2,853	.547

Anmerkungen: P=Pflegekräfte; Ä=Ärzte/Ärztinnen

^a Produkt-Moment-Korrelation (Pearsons r)

^b arithmetisches Mittel (1)=schlechtester Wert, (4)=bester Wert, negativ formulierte Items wurden für die Auswertung umgepolt

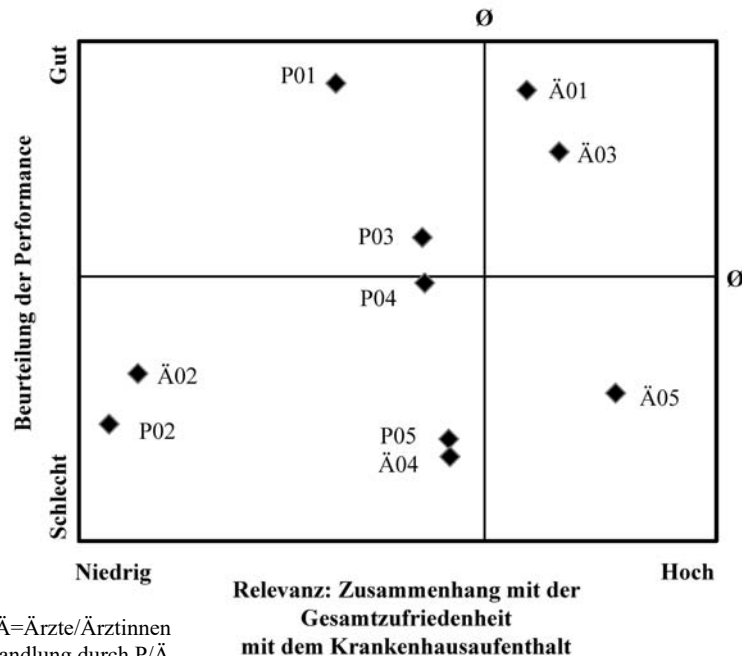
Quelle: Kölner Patientenbefragung, eigene Berechnungen

18 Eine ausführliche Beschreibung der Stichprobe findet sich bei Freise (2003: 39, 161).

19 Eine Liste der zugrunde liegenden Items findet sich in Anhang. Als Antwortskala diente eine vierstufige Zustimmungsskala: stimme überhaupt nicht zu (1) – stimme eher nicht zu (2) – stimme eher zu (3) – stimme voll und ganz zu (4).

20 Es handelt sich um folgendes Item: „Wie zufrieden waren Sie mit Ihrem Krankenhausaufenthalt insgesamt?“ Als Antwortskala diente eine fünfstufige Zufriedenheitsskala: unzufrieden (1) – eher unzufrieden (2) – weder unzufrieden noch zufrieden (3) – eher zufrieden (4) – zufrieden (5).

Abbildung 2: Beispiel – Auszug aus einer Handlungsrelevanzmatrix

**Anmerkungen:**

P=Pflegerkräfte; Ä=Ärzte/Ärztinnen

01 Gerechte Behandlung durch P/Ä

02 P/Ä haben zu häufig gewechselt

03 Fachkompetenz der P/Ä

04 Erreichbarkeit der P/Ä

05 P/Ä führten Gespräche gefühlvoll durch

Quelle: Kölner Patientenbefragung, eigene Darstellung

4.2 Interpretation

Bei der Auswahl von Handlungsfeldern für Verbesserungsmaßnahmen ist es sinnvoll, sich zunächst auf diejenigen Aspekte zu konzentrieren, die eine hohe Relevanz für das Erfolgs- bzw. Zielkriterium besitzen (vgl. Bruhn/Meffert 2002: 66). Der Blick auf die Handlungsrelevanzmatrix richtet sich daher im ersten Schritt auf die Aspekte mit überdurchschnittlicher Wichtigkeit für die Gesamtzufriedenheit mit dem Krankenhausaufenthalt: Dies sind hier drei Leistungsaspekte (Ä01, Ä03, Ä05). Die „gefühlvolle Gesprächsführung durch Ärzte/Ärztinnen“ (Ä05) ist nicht nur überdurchschnittlich relevant, sondern gleichfalls unterdurchschnittlich bewertet. Hier besteht folglich der dringlichste Handlungsbedarf und von der Verbesserung dieses Leistungsaspekts kann ein hoher Effekt auf die Gesamtzufriedenheit mit dem Krankenhausaufenthalt erwartet werden. Die beiden anderen Aspekte „gerechte Behandlung durch Ärzte/Ärztinnen“ und „Fachkompetenz der Ärzte/Ärztinnen“ sind dagegen klar überdurchschnittlich bewertet. Sie sind aktuell Treiber der Gesamtzufriedenheit. Bezüglich dieser beiden Aspekte ist deshalb darauf zu achten,

dass das aktuelle Leistungsniveau erhalten bleibt.

Eine annähernd durchschnittliche, mittlere Wichtigkeit besitzen die Aspekte P03, P04, P05 und Ä04. Die stark unterdurchschnittliche Bewertung der Erreichbarkeit der Ärzte/Ärztinnen (Ä04) und die bemängelte „gefühlvolle Gesprächsführung durch Pflegekräfte“ (P05) machen diese Aspekte zu den nächst dringlichen Ansatzpunkten für Verbesserungsmaßnahmen.

An diesem Beispiel wird ersichtlich, dass die Quadranteneinteilung lediglich als Interpretationshilfe verwendet werden kann. Da sie die mit der Positionierung gegebene Information sehr stark reduziert, lässt die Quadranteneinteilung allein keine differenzierte Interpretation des Ergebnisses zu.

Die mittige Position der Leistungsaspekte Fachkompetenz und Erreichbarkeit der Pflegekräfte (P04, P03) lässt sich inhaltlich besser unter Berücksichtigung der entsprechenden Leistungsthemen bei den Ärzten/Ärztinnen erfassen. Die Erreichbarkeit von Pflegekräften und Ärzten/Ärztinnen (Ä04, P04) besitzt eine ähnliche Relevanz für die Gesamtzufriedenheit. Die Erreichbarkeit der Pflegekräfte wird im Vergleich zu den Ärzten/Ärztinnen jedoch deutlich besser beurteilt. Dieser Vergleich deutet insgesamt nicht auf einen besonderen Handlungsbedarf bzgl. einer Verbesserung der Erreichbarkeit der Pflegekräfte hin. Anders hingegen der Vergleich des Leistungsthemas Fachkompetenz zwischen Pflegekräften und Ärzten/Ärztinnen – diesbezüglich scheinen differenzierte Überlegungen angebracht: Wie man erwarten kann, ist die Fachkompetenz der Pflegekräfte (P03) aus Sicht der Patientinnen und Patienten weniger wichtig für die Gesamtzufriedenheit als die Fachkompetenz der Ärzte/Ärztinnen. Die leicht überdurchschnittliche Beurteilung weist zudem nicht per se auf einen Handlungsbedarf hin. Jedoch kann von einem professionspolitischen Standpunkt aus die schlechtere Beurteilung der Fachkompetenz der Pflegekräfte im Vergleich zu den Ärzten/Ärztinnen kritisch gesehen werden. Dies gilt insbesondere vor dem Hintergrund, dass bzgl. des Leistungsaspektes „gerechte Behandlung“ kein Unterschied zwischen beiden Fachgruppen besteht.

An diesem Beispiel zur Fachkompetenz werden Nutzen und Grenzen des Analysetools Handlungsrelevanzmatrix deutlich. Für die Auswahl von Handlungsfeldern für Verbesserungsmaßnahmen sind in der Praxis mehr Entscheidungsdimensionen relevant, als in der Handlungsrelevanzmatrix berücksichtigt werden können. Die Handlungsrelevanzmatrix ist deshalb auf die Funktion der Entscheidungsunterstützung begrenzt und kann eine Entscheidung der verantwortlichen Akteure – unter sorgfältiger Abwägung aller wichtigen Einflussfaktoren – nicht ersetzen.

In der linken unteren Ecke der Handlungsrelevanzmatrix findet sich mit geringer Relevanz für die Gesamtzufriedenheit der häufige Wechsel von Ärzten/Ärztinnen und Pflegekräften (Ä02, P02). Die Position auf der Wichtigkeits-Achse weist diesen Aspekten – trotz einer schlechten Beurteilung – keinen Handlungsbedarf zu. Die besondere Bedeutung der zweiten Dimension „Wichtigkeit“ für die strategische Auswahl der Handlungsfelder mit Verbesserungsbedarf zeigt sich hier deutlich: Die Aufmerksamkeit und die Ressourcen werden auf diejenigen Entwicklungsfelder gelenkt, die einen hohen Effekt für die Leistungserstellung haben. Die in der praktischen operativen Arbeit häufig präsentere und vertrautere Quali-

tätsdimension (gut – schlecht), erhält als strategisches Entscheidungskriterium nachgeordnete Bedeutung.

Weiterführende Interpretation

Über die konkret handlungsorientierte Interpretation hinaus zeigt die Handlungsrelevanzmatrix, dass die ärztlichen Leistungen im Großen und Ganzen für die Gesamtzufriedenheit wichtiger sind als die pflegerischen Leistungen. Dies ist grundsätzlich nicht überraschend, die Handlungsrelevanzmatrix beschreibt jedoch differenziert, für welche Leistungsaspekte dies gilt (Ä01, Ä03, Ä05) – und für welche nicht (Ä02, Ä04). Die Darstellung der Handlungsrelevanzmatrix ermöglicht es darüber hinaus, unterschiedliche ‚Bündel‘ von Leistungsaspekten in den Blick zu nehmen. Dies ist für die praktische Arbeit wichtig, da Entwicklungsprozesse in der Praxis üblicherweise an bestimmten ‚Themenbereichen‘ ansetzen, z.B. für Berufsgruppen oder bestimmte Prozesse durchgeführt werden. So erhält z.B. die Pflegeleitung – auch unabhängig vom Vergleich zwischen pflegerischen und ärztlichen Leistungsaspekten bzw. dem Gesamtergebnis – Informationen, welche pflegerischen Leistungsaspekte sich besonders für Entwicklungsprozesse innerhalb der Pflege eignen (P05).

4.3 Anmerkungen zur Konstruktion der gezeigten Handlungsrelevanzmatrix

Abhängig von den Adressaten der Untersuchungsergebnisse, kann die Handlungsrelevanzmatrix hinsichtlich des Vereinfachungsgrads der Ergebnisdarstellung variiert werden. Die gezeigte Handlungsrelevanzmatrix stellt einen mittleren Grad der Vereinfachung dar: Die Positionierung der Leistungselemente im Raum sowie die Referenzmarkierungen sind entsprechend den tatsächlichen Messwerten angeordnet. Die Abstände zwischen den Punkten spiegeln somit die Abstände der Messwerte proportional wider. Auf eine Ausweisung der Achsen-Skalierung (wie z.B. bei Matzler/Sauerwein/Heischmidt 2003: 122; Bruhn 2008: 288) wurde jedoch zugunsten einer besseren allgemeinen Verständlichkeit verzichtet. Zudem wird eine Skalierung für die beabsichtigte relative Interpretation der Positionierungen auch nicht benötigt. Eine noch weitergehende Vereinfachung ist beispielsweise in der bei Scharioth, Huber und Jarvis (2004: 154) gezeigten Handlungsrelevanzmatrix realisiert: Die Positionierung der Elemente findet in einem standardisierten Merkmalsraum statt, der immer quadratisch ist und bei welchem sich das Achsenkreuz der Quadrantengrenzen immer in der Mitte befindet (vgl. auch Jarvis 2008: 6f.).²¹

21 Eine solche Standardisierung des Merkmalsraumes stellt Jarvis (2008: 6f.) zufolge eine Vereinfachung für das Lesen der Handlungsrelevanzmatrix und eine konsequent weitergeführte Umsetzung des relativen Interpretationsansatzes dar. Wird der Merkmalsraum nicht standardisiert, richtet sich die Position des Achsenkreuzes immer nach den aktuellen Daten, wodurch das Aussehen der Handlungsrelevanzmatrix variiert. Ein Vorteil des unstandardisierten Merkmalsraums besteht jedoch darin, dass sich die Positionierung direkt aus den statistischen Maßen ergibt, was wiederum der Nachvollziehbarkeit der Konstruktion der Handlungsrelevanzmatrix entgegen kommt. Die relative Interpretation der Positionierung ist zudem auch im unstandardisierten Merkmalsraum möglich.

5. Diskussion

Damit Handlungsrelevanzmatrizen einen nutzenstiftenden Beitrag in der Evaluationsforschung leisten können, müssen sie zwei zentrale Anforderungen erfüllen: Zum einen müssen die generierten Ergebnisse eine hinreichende Qualität und Aussagekraft besitzen, um als Basis für Entscheidungen dienen zu können (methodische Qualität) und zum anderen muss das Evaluationsprojekt so gestaltet sein, dass die Ergebnisse in praktischen Arbeitszusammenhängen verwendet werden können (Nützlichkeit) (vgl. z.B. Bortz/Döring 2005: 102ff.; Meyer 2007b: 143). Die Evaluationsstandards der DeGEval (Gesellschaft für Evaluation e.V.) nennen neben diesen beiden Aspekten der (methodischen) „Genauigkeit“ und „Nützlichkeit“ als zwei weitere zentrale Qualitätskriterien: die „Durchführbarkeit“ und die „Korrektheit“. Der Standard der Durchführbarkeit hebt darauf ab, dass die Projektkonzeption den praktischen Kontext berücksichtigt und das Evaluationsprojekt vor diesem Hintergrund realistisch geplant wird. Der Standard der „Korrektheit“ zielt darauf ab, dass die Evaluation rechtlich und ethisch korrekt sowie fair erfolgt (vgl. Widmer 2006; DeGEval 2008). Die Betrachtung der beiden letztgenannten Standards findet sich nachfolgend im Rahmen der Diskussion von methodischer Qualität und Nützlichkeit.

5.1 Methodische Qualität

Die methodische Qualität der Evaluation bzw. – daraus resultierend – die Qualität der Ergebnisse ist für eine steuerungs- und entwicklungsorientierte Evaluation von besonderer Bedeutung. Irreführende Ergebnisse können zu falschen Schlussfolgerungen führen, die sowohl kostspielige Fehlinvestitionen nach sich ziehen als auch weitreichende, nachteilige Folgen für die betroffenen Personen haben können (vgl. Kromrey 2007: 115; Freise 2003: 76; Töpfer/Mann 1996; Menold 2007: 26). Die zwei zentralen methodischen Qualitätskomponenten einer befragungsbasierten Evaluation sind einerseits die (realisierte) Stichprobe und andererseits der Fragebogen bzw. die verwendeten Messinstrumente.

5.1.1 Stichprobenprobleme

In praktischen Arbeitszusammenhängen ist es häufig schwierig, eine Zufallsstichprobe zu realisieren. Zudem sind niedrige Rücklaufquoten kein seltenes Problem und es ergeben sich häufig bereits aus dem Befragungsverlauf deutliche Hinweise, dass bestimmte Personengruppen sich nur unterdurchschnittlich an der Befragung beteiligt haben. Dies bedeutet, dass die realisierte Stichprobe verzerrt ist. Barkley und Furse (1996) zeigen anhand einer Patientenstichprobe, dass verzerrte Stichproben zu falschen Handlungsempfehlungen führen können, wenn lediglich auf einfache, verzerrungssensible statistische Maße, wie z.B. Mittelwerte und Häufigkeiten zurückgegriffen wird.

Die Analyse anhand der Handlungsrelevanzmatrix verwendet zusätzlich zur einfachen, mittel- oder anteilswertbasierten Performancemessung die zusammenhangsanalytisch ermittelte Wichtigkeit. Diese Art der Ergebnisanalyse hat den Vor-

teil, vergleichsweise wenig anfällig für Stichprobenverzerrung zu sein. Die Schätzung von Korrelationen ist gegenüber systematischen Stichprobenfehlern deutlich robuster als die Schätzung von Anteilswerten und Mittelwerten (vgl. Diekmann 1998: 364) und es wird eine zweite, zusätzliche Dimension für die Ergebnisinterpretation herangezogen.

5.1.2 Single-Item- versus Multiple-Item-Messung

Das zentrale konzeptionelle Element der Handlungsrelevanzmatrix ist die Messung der einzelnen Leistungselemente auf Basis von Einzelitems (Single-Item-Messung). Diese Vorgehensweise kommt zum Einsatz, da sie sehr kostensparend ist, weil bei einer gegebenen Fragebogenlänge eine große Anzahl von Einzelaspekten abgefragt werden kann. Auf diese Weise wird ein detailliertes Bild, basierend auf vielen unterschiedlichen Aspekten des Dienstleistungsgeschehens, erhoben (multiattributive Messung) (vgl. Drolet/Morrison 2001: 196; Nagy 2002: 79).

Das Konzept der Single-Item-Messung widerspricht jedoch der gängigen Auffassung in der Methodenliteratur, die für eine valide Messung eine Mehr-Item-Messung fordert (vgl. z.B. Wanous/Reichers/Hudy 1997: 247). Die drei zentralen Argumente für die Mehr-Item-Messung bestehen darin, dass erstens durch die Verwendung mehrerer Items die Fehlerkomponente der Messung verringert wird (vgl. Bühner 2006: 25, 29; Peter 1979: 7; Churchill 1979: 66; Matz 2007: 27), zweitens für Mehr-Item-Messinstrumente etablierte statistische Maße zur Bestimmung der Qualität bzw. Güte der Messung existieren, was bei Ein-Item-Instrumenten nicht der Fall ist (vgl. Drolet/Morrison 2001: 196; Menold 2007: 32), und drittens Mehr-Item-Instrumente es ermöglichen, die verschiedenen Niveaus der Ausprägung eines Konstrukts genauer zu bestimmen (vgl. Netemeyer/Bearden/Sharma 2003: 5; Meyer 2007a: 209; Churchill 1979: 66).²²

22 Anzumerken ist, dass sich nicht für alle Mehr-Item-Messinstrumente statistische Gütemaße bestimmen lassen. Dies ist nur für Skalen möglich, nicht aber für Indizes. Die Begriffe „Index“ und „Skala“ werden in den Sozialwissenschaften uneinheitlich verwendet, es handelt sich jedoch um unterschiedliche Messkonzepte (vgl. Greving 2007: 66; Schnell/Hill/Esser 1999: 166). Indizes sind Mehr-Item-Messinstrumente, die mehrere Dimensionen beinhalten, Skalen sind dagegen Mehr-Item-Messinstrumente, die nur eine Dimension messen sollen. Skalen können daher als Spezialfälle von Indizes angesehen werden. Im Gegensatz zu Skalen, für deren Konstruktion und Validierung elaborierte statistische Vorgehensweisen und Gütekriterien existieren, erfolgt die Konstruktion eines Index vorwiegend auf theoretischer Basis, weitgehend unabhängig von den Daten (vgl. Schnell/Hill/Esser 1999: 168). Das Kriterium für die Qualität von Indizes ist inhaltlicher Art und besteht darin, dass alle für das Messkonstrukt relevanten Dimensionen im Index enthalten sind. Rückschlüsse auf die Güte von Indizes gibt somit allenfalls die theoretische Fruchtbarkeit der Analyse (vgl. Schnell/Hill/Esser 1999: 168). Ein Index wird dann verwendet, wenn ein einzelner Indikator die interessierende Dimension nicht genau misst oder sozialwissenschaftliche Begriffe mehrere Dimensionen ansprechen (vgl. Schnell/Hill/Esser 1999: 168). Mit der Verknüpfung verschiedener Indikatoren zu einem Index (bzw. der Index-Maßskala) wird das Ziel verfolgt, komplexe Sachverhalte möglichst einfach darzustellen. Es wird eine Kennzahl (One-Number-Tool) konstruiert, mit welcher ein komplexer Sachverhalt anhand eines einzelnen Zahlenwertes auf einer einzigen Index-Skala bewertet wird (öffentlich bekannte Indizes sind z.B. der Konsumklimaindex der GfK oder der Geschäftsklimaindex des Ifo-Instituts). Ein Index erleichtert auf diese Weise ganz wesentlich die Arbeit in Entscheidungsprozessen (vgl. Meyer 2007a: 214ff.), weshalb sie in den Sozial- und Wirtschaftswissenschaften häufig Anwendung finden.

Vergleichende Untersuchungen zur Qualität von Single- und Multiple-Item-Messung gibt es nur wenige. Nagy (2002) untersuchte beide Vorgehensweisen im Rahmen einer Studie zur Arbeitszufriedenheit. Einige seiner Ein-Item-Instrumente wiesen sogar eine höhere Vorhersagevalidität (prädiktive Validität) als die entsprechenden Skalen auf. Er kommt daher zum Schluss, dass für Fragen der Arbeitszufriedenheit die Single-Item-Messung vorzuziehen sei – auch wenn in diesem Fall keine statistischen Gütemaße vorhanden sind. Drolet und Morrison (2001) verglichen die beiden Vorgehensweisen im Rahmen der Messung von Dienstleistungsqualität. Die Studie zeigt, dass die Verringerung des Messfehlers durch ein zweites oder drittes Item meist sehr gering ausfällt und wenig zusätzliche Information erbringt. Ferner bestünde bei der Mehr-Item-Messung die Gefahr unerwünschten Antwortverhaltens, welches zwar die tatsächliche Validität und Reliabilität verringere, sich jedoch nicht in den klassischen Gütemaßen für die Reliabilität einer Skala niederschlägt.²³ Die Autoren empfehlen daher insbesondere für die Forschung im Bereich der Dienstleistungen die Verwendung der Single-Item-Messung.

Der bei Drolet und Morrison (2001) angesprochene Verwendungszusammenhang als Kriterium für die Entscheidung über die Gestaltung der Messinstrumente wird bislang kaum diskutiert. Bei Bortz/Döring (2005: 253) findet sich diesbezüglich der Hinweis, dass Fragebögen zur Erfassung von Persönlichkeitsmerkmalen anders zu konzipieren sind als Fragebögen, bei welchen es um Verhaltensweisen oder die Beschreibung und Bewertung von Sachverhalten geht. Die Forderung nach Multiple-Item-Messung, die einen wesentlichen Ursprung in der psychologischen oder diagnostischen Forschung hat, stellt – den beiden angeführten Studien zufolge (Drolet/Morrison 2001; Nagy 2002) – bei der Evaluation von Dienstleistungen eher ein ‚Overengineering‘ dar: die durch die Messung mit einer Skala aus mehreren Items erzielte höhere Genauigkeit erbringt für Dienstleistungsfragestellungen keinen relevanten Zusatznutzen, geht aber mit dem Nachteil einher, dass es bei begrenztem Fragebogenumfang kaum möglich ist, eine große, aussagekräftige Anzahl von Einzelaspekten des Dienstleistungsgeschehens zu erheben.²⁴ Einen gewichtigen Nachteil hat die Single-Item-Messung jedoch in jedem Fall: Es fehlen standardisierte statistische Gütekriterien (vgl. Nagy 2002: 84).

23 Drolet und Morrison (2001: 200) zeigen, dass die Verwendung mehrerer Items zur Messung eines Sachverhalts eher zu gedankenlosem Ankreuzen und somit zu Validitätsproblemen aufgrund des Antwortverhaltens führt. Die zusätzliche Fragebogenlänge und Zeit, die durch die Mehr-Item-Messung benötigt werden, seien diesem Validitätsproblem gegenüber sogar nur sekundär.

24 Für die psychologische und pädagogische Forschung, welchen die etablierten Techniken der Skalenkonstruktion und der statistischen Gütebestimmung von Skalen wesentlich zuzuordnen sind (vgl. z.B. Meyer 2007a: 219; Menold 2007: 32ff.), findet die rigide Forderung nach Multiple-Item-Messung und statistischen, belastbaren Gütekriterien ihre Berechtigung ebenfalls im Verwendungszusammenhang. Die Skalen werden teilweise zu diagnostischen Zwecken eingesetzt, d.h. anhand des Testergebnisses wird eine einzelne befragte Person z.B. einer bestimmten Behandlung zugeführt oder – bei Leistungs- und Berufseignungstests – von bestimmten beruflichen Möglichkeiten ausgeschlossen. Sehr hohe methodische Anforderungen sind bei einer derartigen, individualdiagnostischen Verwendung unabdingbar (vgl. Bortz/Döring 2005: 188ff., 253f.). Die Ergebnisse von Dienstleistungsstudien beziehen sich dagegen fast ausschließlich nicht auf einzelne Befragte, sondern auf Gruppen von Befragten (z.B. Altersgruppen, Einkommensgruppen etc.).

Die Frage nach Ein-Item oder Mehr-Item-Messung ist für die beiden Erhebungselemente der Handlungsrelevanzmatrix, die Leistungsaspekte und das Erfolgskriterium, differenziert zu diskutieren. Wie dargestellt steht bezüglich der Erhebung des Leistungsgeschehens vor allem eine große Anzahl von Einzelaspekten im Vordergrund. Hingegen finden sich gute Argumente für die Messung des Erfolgskriteriums in Form eines Index bzw. einer mehr-item-basierten Kennziffer.²⁵ Gerade für soziale Dienstleistungen existieren sehr unterschiedliche und komplexe Erfolgsdimensionen (z.B. Kundenzufriedenheit, Wirksamkeit/Nutzen, Effizienz, Menschlichkeit/Gerechtigkeit). Die Messwerte zu verschiedenen einzelnen Erfolgskriterien können im Rahmen der Indexbildung zu einem gesamten, abstrakten Index-Messwert zusammengefasst werden. Durch die Verwendung eines Index wird auf diese Weise ein komplexer, mehrdimensionaler Sachverhalt – wie „Erfolg“ – einfach und übersichtlich in einer Skala dargestellt und durch die Indexierung wird eine höhere Varianz bei der Messung des Erfolgskriteriums erreicht (vgl. Meyer 2007a: 214ff.; Churchill 1979: 66).²⁶

5.2 Praktische Nützlichkeit

Ziel der Verwendung empirischer Methoden in praktischen Entwicklungszusammenhängen ist es, situationsadäquates Handlungswissen zu generieren und dieses anschließend in die Praxis zu übertragen, um die Praxis bei der Umsetzung ihrer Aufgaben zu beraten (vgl. z.B. Bortz/Döring 2005: 103ff.; Dickopf/Pies 2004; Kromrey 2007: 114). Neben das klassisch-wissenschaftliche Erfolgskriterium des Erkenntnisgewinns tritt für praxisorientierte empirische Vorgehensweisen das Kriterium der praktischen Nützlichkeit. Nützlich sind Evaluationsprojekte für die Praxis insbesondere dann, wenn sie

- a) *anwendungsbezogen* sind, d.h. wenn praktische Fragestellungen mit einer angemessenen Methodik in ein Forschungsdesign überführt werden und dabei mit engem und nachvollziehbarem Bezug zum Untersuchungsgegenstand (der Dienstleistung) gearbeitet wird;
- b) *entwicklungs- und umsetzungsbezogen* sind, d.h. wenn die Evaluation und ihre Ergebnisse Veränderungen (bzw. Verbesserungen) in der Praxis ermöglichen oder fördern (vgl. z.B. Müller 2009: 146ff.).

25 Für eine intensive Diskussion der Frage der Messung des Erfolgskriteriums als Index oder Einzelitem sei auf die Literatur zur Kundenzufriedenheitsmessung verwiesen. An dieser Stelle seien als besonders interessant die beiden Texte von Reichheld (2003) und Bachmann (2007) genannt, die die weit verbreiteten Vorbehalte gegenüber Ein-Item-Messinstrumenten relativieren.

26 Wird als Erfolgskriterium eine Maßzahl (KPI) verwendet, die auch in andere Managementinstrumente eingeht, z.B. in die Balanced Scorecard, können auf diese Weise die strategische und operative Steuerungsebene auch in der Methodik verknüpft werden. Handlungsrelevanzmatrizen werden damit ein integraler Bestandteil eines umfassenden Managementinformationssystem (vgl. Huber 2008: XI; Huber/Scharioth 2003; Hermann 2006: 6ff.).

5.2.1 Multiperspektivisches Forschungsdesign

Mittlerweile wird an vielen Stellen in der Evaluationsliteratur gefordert, dass angemessene Evaluationsdesigns für soziale Dienstleistungen multiperspektivisch angelegt sein sollen (vgl. Koch/Fertsch-Röver 2009: 123). Soziale Dienstleistungen werden zu einem großen Teil im Rahmen eines Ko-Produktionsverhältnisses erstellt, d.h. die Leistungserstellung ist wesentlich vom Zusammenwirken aller Beteiligten abhängig. Die verschiedenen Beteiligten(-gruppen) haben dabei aufgrund ihrer unterschiedlichen Rollen und Funktionen verschiedene Sichtweisen auf und Interessen an der Dienstleistung und richten ihr Handeln danach aus (vgl. z.B. Wolff/Scheffer 2002: 332). Die Berücksichtigung der unterschiedlichen Wahrnehmungen und Interessenlagen der verschiedenen Beteiligtegruppen ist somit für das Verständnis des Dienstleistungsgeschehens und der Resultate essentiell (vgl. Lüders/Haubrich 2002: 326). Ein multiperspektivischer Zugang wird der Unterschiedlichkeit der möglichen Referenzsysteme dabei am ehesten gerecht (vgl. Schiepek/Bauer 1998: 22).

Die methodische Umsetzung von Multiperspektivität gestaltet sich jedoch schwierig und resultiert nicht selten in komplexen, vielschichtigen und schwer auswertbaren Erhebungsdesigns (vgl. z.B. Müller 2009; Schuster 2007: 172; Koch/Fertsch-Röver 2009: 123). Handlungsrelevanzmatrizen können diesbezüglich eine Lösung darstellen, da sie ein einheitliches Auswertungs- (und Erhebungs-)format für die verschiedenen Perspektiven aller Beteiligtegruppen bieten (vgl. Scharioth/Huber 2003: 6; Hermann 2006: 6). Die Methodik ist gleichzeitig inhaltlich sehr flexibel. So kann sowohl das Set der Leistungsaspekte für die verschiedenen Stakeholdergruppen unterschiedlich gestaltet werden, als auch die Erfolgsdimension jeweils entsprechend der Interessenlage bzw. dem Bezug der unterschiedlichen Stakeholdergruppen zum Evaluationsgegenstand definiert werden (vgl. Töpfer 1996: 238f.; Hermann 2006: 6). Handlungsrelevanzmatrizen liefern schließlich in übersichtlicher, einheitlicher Form für jede Stakeholdergruppe eine differenzierte Bewertung des Dienstleistungsgeschehens, Erkenntnisse zur Relevanz einzelner Leistungsaspekte sowie eine Priorisierung von Veränderungsbedarfen. Auf dieser einheitlichen Grundlage lassen sich die verschiedenen Perspektiven einfach aufeinander beziehen und inhaltlich vergleichen, ohne dass die Übersicht verloren geht. Es entsteht eine auf die Ansatzpunkte für Veränderungsbedarf zugespitzte ‚Rundumsicht‘ über das Dienstleistungsgeschehen, die direkt in Entscheidungsfindungsprozesse einfließen kann. Durch Kombination mit anderen Analyseinstrumenten können Handlungsrelevanzmatrizen darüber hinaus zu einem differenzierten Stakeholder-Management-Systems ausgebaut werden (vgl. Huber 2008).

5.2.2 Passgenauigkeit der Fragen

Je besser der Fragebogen die nach Branche und Dienstleistung variierenden Leistungs- und Qualitätsaspekte reflektiert und auf die spezifischen Untersuchungsfragenstellungen im konkreten Praxiszusammenhang eingeht, desto fruchtbarer sind die Ergebnisse anschließend für die Entwicklung der Praxis (vgl. Thayer/Fine 2001;

Bruhn 2008: 288). Eine Formulierung passgenauer, detaillierter und erschöpfender Fragen für die Dienstleistungssituation in einem konkreten praktischen Zusammenhang ist dabei realistischerweise nicht durch das Evaluationspersonal allein leistbar, sondern nur durch die Einbeziehung der Praktiker/-innen vor Ort bzw. aller relevanten Stakeholdergruppen in die Fragebogenkonstruktion zu erreichen (vgl. Bortz/Döring 2005: 108ff.). Dadurch, dass die Beurteilung der Leistungselemente in Handlungsrelevanzmatrizen auf einfachen Ein-Item-Messinstrumenten aufbaut und für alle Elemente dieselbe, einheitliche Beurteilungsskala verwendet wird, bestehen methodisch keine nennenswerten Hürden für die Beteiligung von Praktikern/Praktikerinnen an der Item-Formulierung für die Leistungsaspekte. Diese Vorgehensweise ermöglicht ferner auch die einfache Aufnahme neuer Aspekte in den Fragebogen, insbesondere auch solcher, für die keine validierten Skalen existieren. Bei Wiederholungsbefragungen lassen sich darüber hinaus auf einfache Weise Fragen verändern oder neue Fragen hinzufügen (vgl. Nagy 2002: 79). Ohne die Begleitung von Experten/Expertinnen sind Handlungsrelevanzmatrizen in der Regel dennoch nicht einsetzbar. Insbesondere für die Konzeption einer validen Messung der Erfolgsdimension, die für die Bestimmung der Relevanz zentral ist, sind in der Regel tiefere methodische und inhaltliche Kenntnisse notwendig. Dies gilt insbesondere dann, wenn das Erfolgskriterium Benchmarking-Analysen ermöglichen und daher auch situationsübergreifend einsetzbar sein soll.

5.2.3 Verständlichkeit und Akzeptanz

Die Umsetzung von Evaluationsergebnissen in Strukturen und Handlungsrouinen durch die Praxis hängt wesentlich davon ab, dass die Praktiker/-innen die Ergebnisse und Empfehlungen sowohl verstehen als auch akzeptieren (vgl. Dickopf/Pies 2004: 36; Schrappner 2004: 19; Bortz/Döring 2005: 102f.). Für die Akzeptanz von Handlungsempfehlungen ist von besonderer Bedeutung, dass deren Ableitung aus den Ergebnissen für Praktiker/-innen nachvollziehbar wird. Die grafikbasierte anstelle einer zahlenbasierten Ergebnisdarstellung trägt diesem Bedarf Rechnung. ‚Gute‘ und ‚schlechte‘, ‚wichtige‘ und ‚unwichtige‘ Ergebnisse sowie die Handlungsdringlichkeiten ergeben sich in der Handlungsrelevanzmatrix unmittelbar aus der Ergebnisdarstellung. Die Ableitung der Handlungsempfehlungen aus den Ergebnissen lässt sich auch von methodischen Laien leicht erschließen. Dies geht mit einem weiteren Vorteil einher: mit den Ergebnissen kann direkt, d.h. ohne weitere Aufbereitung, in praktischen Arbeitszusammenhängen gearbeitet werden. Beim erstmaligen Einsatz von Handlungsrelevanzmatrizen ist die Einführung dieses Tools durch einen geschulten Moderator bzw. eine geschulte Moderatorin sicherlich empfehlenswert, aber selbst dies ist Bauer (1996: 126) zufolge nicht notwendig.

5.2.4 Entwicklungs- und Umsetzungsorientierung

Zentrales Kriterium für den Nutzen von Handlungsrelevanzmatrizen als Tool der Evaluation ist die Eignung, einen spezifischen Beitrag zur Entwicklung und Verbesserung der Praxis zu leisten. Durch die doppelte Orientierung an der Perfor-

mance und der Relevanz bringen Handlungsrelevanzmatrizen die Evaluationsergebnisse auf den für die Praxisentwicklung relevanten Punkt. Die gesammelten Informationen werden so dargestellt, dass die personellen und finanziellen Ressourcen bestmöglich auf diejenigen wenigen Entwicklungsaspekte konzentriert werden können, die die größte Wirkung haben (vgl. Slack 1994: 73; Hermann 2006: 1; Sampson/Showalter 1999: 2; Töpfer 1996: 255).

Darüber hinaus haben sich Handlungsrelevanzmatrizen in ihrem bisherigen, vornehmlich privatwirtschaftlichen Einsatzbereich, nicht nur als besonders praxisorientierte Auswertungsvariante sondern auch als Diskurs-Instrument zur Förderung und Strukturierung der Kommunikation in Entwicklungsprozessen bewährt (vgl. Bauer 1996: 126; Slack 1994: 74). Bauer (1996: 126) beschreibt den Nutzen des von ihm verwendeten Handlungsrelevanzmatrix-Verfahrens wie folgt: „(...) [es] ist vorrangig als strukturierte Kommunikationsmethode gedacht. Es ist einfach zu verstehen und anzuwenden. ... Es werden fast ausnahmslos und von Anfang an gute Ergebnisse erzielt.“ In Bezug auf die Zusammenarbeit zwischen Abteilungen im Unternehmen fügt er hinzu: „Teilweise werden jahrelang gestörte und festgefahrene Beziehungen sprunghaft verbessert“ (ebd.).²⁷ Zum Diskurs-Instrument für Verbesserungsprozesse qualifiziert sich die Handlungsrelevanzmatrix insbesondere durch den in der Positionierung verwirklichten relativen Interpretationsansatz. Unabhängig vom absoluten Qualitätsniveau werden immer sowohl Schwächen als auch Stärken der Leistungserbringung thematisiert. Dies entspricht einerseits dem Prinzip der kontinuierlichen Qualitätsverbesserung (vgl. Jarvis 2008: 6f.) und ermöglicht andererseits eine ausgewogene, faire Evaluation und Diskussion. Gerade diese Eigenschaft ist für die Entwicklung von sozialen, ko-produktiven Dienstleistungsprozessen besonders bedeutsam: Das Lernen des Fachpersonals und dessen verändertes Handeln im Ko-Produktionsprozess sind zentrale Faktoren für die Verbesserung der Dienstleistung.

Handlungsrelevanzmatrizen können sowohl für die Steuerung sozialer Dienstleistungen auf der Managementebene als auch für Reflexionsprozesse auf der Ebene der professionell Tätigen verwendet werden. Die verständliche Darstellungsform von Zusammenhängen zwischen einzelnen Leistungselementen und ihrer Erfolgswirksamkeit schlägt eine Brücke zwischen einem „wohlfahrtsmanageriellen“ Steuerungsansatz und einem reflexionsorientierten Professionalitätsverständnis (vgl. Otto 2007: 89ff.). Die breite Verständlichkeit der Handlungsrelevanzmatrix erfüllt damit eine wichtige Voraussetzung dafür, dass Evaluations- und Wirkungsforschung in partizipative Entwicklungsprozesse münden kann und nicht in die Situation gerät, zu einer Deprofessionalisierung beizutragen (vgl. Otto 2007: 90ff.).

5.2.5 Grenzen der Handlungsrelevanzmatrix

Handlungsrelevanzmatrizen sind für Entwicklungsfragestellungen in der Praxis konstruiert und basieren auf einer Kombination vergleichsweise einfacher methodi-

²⁷ Der Autor kann sich dieser Bewertung aus eigener, mehrjähriger praktischer Erfahrung in der Anwendung von Handlungsrelevanzmatrizen als Evaluations- und Beratungstool anschließen.

scher Instrumente. Die Grenzen der Handlungsrelevanzmatrix verlaufen dort, wo Studien einen grundlagenorientierten Charakter annehmen. So sind z.B. für den Test von Wirkungsmodellen andere Forschungsdesigns und andere statistische Analyseverfahren zu wählen. Kritisiert wird an Handlungsrelevanzmatrizen in diesem Zusammenhang insbesondere die einfache, korrelative Messung der Relevanz von Leistungsaspekten für das Erfolgskriterium. Diese Kritik relativiert Jarvis (2008: 6) dahin gehend, dass es für den Verwendungszweck viel bedeutsamer sei, bei Verwendung einfacher Zusammenhangsmaße eine große Anzahl von Aspekten in der Analyse berücksichtigen zu können.²⁸ Vor dem Hintergrund der Zielsetzung der Praxisentwicklung ist die methodische Schlichtheit daher nicht als methodischer Mangel zu sehen, sondern als Konstruktionsprinzip, um die Ergebnisse verständlich und nachvollziehbar zu halten sowie dem Verwendungszweck gerecht zu werden.

Ein weiterer Kritikpunkt an Handlungsrelevanzmatrizen und ähnlichen Analyseinstrumenten (z.B. Portfolioanalyse) betrifft das grundlegende Konstruktionsprinzip der Reduktion auf zwei zentrale Dimensionen (vgl. Kortendieck 2008: 63; Esch/Herrmann/Sattler 2006: 178; Pracht/Bachert 2005: 82). Die Handlungsrelevanzmatrix reduziert die Analyse des Leistungsgeschehens auf die Performance und die Relevanz für *ein* definiertes Erfolgskriterium.²⁹ Einige Facetten der Leistungserstellung sowie andere als das gewählte Erfolgskriterium werden somit ausgeblendet. Diese Reduktion ist gleichzeitig die Grundlage für die zentralen Vorteile des Instrumentes: die einfache Darstellung, die Klarheit der Priorisierung und die leicht verständliche Ableitung von Handlungsempfehlungen. Für die Verwendung von Handlungsrelevanzmatrizen in Entwicklungs- und Steuerungsprozessen ist daher zu betonen, dass sie Managemententscheidungen nicht ersetzen, sondern lediglich unterstützen können. Dies ist jedoch kein besonderes Charakteristikum der Handlungsrelevanzmatrix, sondern gilt für Evaluationsforschung generell (vgl. Meyer 2007a: 220).

28 Die Verwendung von Korrelationen zur Relevanzmessung wird insbesondere dahin gehend kritisiert, dass erstens keine Kausalaussagen und zweitens keine Aussagen zur Richtung eines Ursache-Wirkungsverhältnisses gemacht werden können; drittens bleiben Interkorrelationen und Kolinearität unberücksichtigt. Wie Jarvis (2008: 6) ausführt, ist der zentrale Vorteil der Verwendung von Korrelationen weitaus bedeutsamer: Es können problemlos 40 bis 60 Items mittels Korrelation analysiert werden. Eine Regression erreicht ihr sinnvolles Maximum bei 10 bis 15 Variablen, so dass bei einer höheren Anzahl von Variablen eine Vorgehensweise mit zwei Arbeitsschritten notwendig wird, d.h. eine vorherige Verdichtung auf Faktoren durchgeführt werden muss. Die Verwendung von Regressionen beschreibt auch Stauss (1999: 14) kritisch: Eine relativ einfache regressionsanalytische Vorgehensweise ist die Verwendung der Beta-Werte als Bedeutungsgewichte. Die methodisch korrektere, aber aufwändigere Vorgehensweise besteht in der bereits beschriebenen Kopplung einer Regressionsanalyse mit einer vorherigen Faktorenanalyse, die Einzelattribute zu wenigen Faktoren zusammenfasst. Letzteres Vorgehen verhindert jedoch wiederum die eigentliche Zielsetzung der Ausweisung von Wichtigkeiten für die Einzelattribute.

29 In die Handlungsrelevanzmatrix kann jeweils nur ein Erfolgskriterium als abhängige Variable einfließen. Verschiedene Dimensionen des ‚Erfolgs‘ einer Dienstleistung können entweder dadurch berücksichtigt werden, dass als Erfolgskriterium ein Index verwendet wird, der alle relevanten Erfolgsdimensionen einschließt (vgl. Fußnote 22). Alternativ können mehrere Erfolgskriterien – mehrere Einzelindikatoren oder mehrere Erfolgsindizes – im Fragebogen separat erhoben und für diese jeweils eigene Handlungsrelevanzmatrizen erstellt werden, die sich dann hinsichtlich der inhaltlichen Aussagen vergleichen lassen.

6. Fazit

Handlungsrelevanzmatrizen sind eine spezielle Form der Aufbereitung von Befragungsdaten und stellen ein praxistaugliches Tool zur Identifizierung und Priorisierung von Handlungsfeldern mit Verbesserungsbedarf dar. Im privatwirtschaftlichen Bereich kommen sie regelmäßig zur Anwendung. Zur Evaluation und Verbesserung von sozialen Dienstleistungen werden sie bislang jedoch wenig genutzt, obwohl sie gerade für die komplexen Fragestellungen in diesem Feld eine besonders geeignete Vorgehensweise darstellen.

Handlungsrelevanzmatrizen ermöglichen es, im Rahmen von Befragungen detailliert unterschiedliche Aspekte des Leistungsgeschehens hinsichtlich des Leistungsniveaus (Performance) und hinsichtlich des Beitrags zum Erfolg bzw. der Wirksamkeit der Dienstleistung (Relevanz) zu analysieren. Die Ergebnisse werden grafisch, in leicht verständlicher Form dargestellt, so dass auch methodische Laien nachvollziehen können, wie sich Ansatzpunkte für Verbesserungen und Handlungsempfehlungen aus den Ergebnissen ableiten. Die gute Verständlichkeit qualifiziert Handlungsrelevanzmatrizen sowohl für die Verwendung als Instrument der Steuerung durch das Management als auch als gesamtbetriebliches Kommunikationsmittel zur Konzeption und Abstimmung von Entwicklungsprozessen.

Bisherige Evaluationsansätze im Bereich der sozialen Dienstleistungen stellen häufig die Analyse der Wirksamkeit in den Vordergrund. Für die Praxis schließt sich an die Wirksamkeitsanalyse in der Regel die Frage nach notwendigen Veränderungen an. Eine Integration des Bausteins Handlungsrelevanzmatrix in die gängige Evaluationspraxis könnte auf sehr effiziente und effektive Weise wertvolle Zusatzinformationen für die konkrete Verbesserung sozialer Dienstleistungen bereitstellen.

7. Literatur

- Bachmann, Anne (2007): Subjektive versus objektive Erfolgsmaße. In: Albers, Sönke/Klapper, Daniel/Konradt, Udo/Walter, Achim/Wolf, Joachim (Hg.): Methodik der empirischen Forschung. Wiesbaden: Gabler, S. 89-102.
- Barkley, William M./Furse, David H. (1996): Changing priorities for improvement. The impact of low response rates in patient satisfaction. In: The Joint Commission Journal on Quality Improvement, 22 (6), S. 427-433.
- Bauer, Ewald (1996): Kundenzufriedenheits-Portfolio: Wie interne und externe Kundenerwartungen und Kundenzufriedenheit strukturiert verbessert werden. In: Töpfer, Armin (Hg.): Kundenzufriedenheit messen und steigern. Neuwied u.a.: Luchterhand, S. 115-146.
- Becker, Helmut (2009): Schema einer Handlungsrelevanzmatrix. Verfügbar unter: <http://www.helmutbecker.de/pages/angebot/marktforschung/analysen/matrix/> [01.07.2009].
- Bewyl, Wolfgang/Speer, Sandra/Kehr, Jochen (2004): Wirkungsorientierte Evaluation im Rahmen der Armuts- und Reichtumsberichterstattung – Perspektivstudie. Bundesministerium für Gesundheit und Soziale Sicherung (BMGS). Verfügbar unter: http://www.bmas.de/portal/1850/property=pdf/wirkungsorientierte_evaluation_im_rahmen_der_armutsberichterstattung.pdf [15.01.2011].
- Bortz, Jürgen/Döring, Nicola (2005): Forschungsmethoden und Evaluation. für Human- und Sozialwissenschaftler. Berlin, Heidelberg: Springer-Verlag.
- Bruhn, Manfred/Meffert, Heribert (2002): Exzellenz im Dienstleistungsmarketing. Wiesbaden: Gabler.

- Bruhn, Manfred (2008): Qualitätsmanagement für Dienstleistungen. Berlin, Heidelberg: Springer-Verlag.
- Bühner, Markus (2006): Einführung in die Test- und Fragebogenkonstruktion. München: Pearson Studium.
- Churchill, Gilbert A. (1979): A paradigm for developing better measures of marketing constructs. In: *Journal of Marketing Research*, 16 (February), S. 64-73.
- Cronbach, Lee J. (1982): *Designing Evaluation of Educational and Social Programs*. San Francisco: Jossey-Bass.
- DeGEval – Gesellschaft für Evaluation e.V. (2008): Standards für Evaluation. Verfügbar unter: <http://www.degeval.de/calimero/tools/proxy.php?id=19076> [26.07.2010].
- Dichtl, Erwin/Müller, Stefan (1986): Anspruchsinflation und Nivellierungstendenz als messtechnische Probleme in der Absatzforschung. In: *Marketing ZFP*, 8 (4), S. 233-236.
- Dickopf, Christa/Pies, Silke (2004): Zwischen Erkenntnisgewinn und Praxisentwicklung. Merkmale und Spannungsfelder sozialpädagogischer Forschungsprojekte. In: Schrapper, Christian (Hg.): *Sozialpädagogische Forschungspraxis. Positionen, Projekte, Perspektiven*. Weinheim: Juventa, S. 29-37.
- Diekmann, Andreas (1998): *Empirische Sozialforschung. Grundlagen, Methoden, Anwendungen*. Reinbek: Rowohlt Tb.
- Drolet, Aimee L./Morrison, Donald G. (2001): Do we really need multiple-item measures in service research? In: *Journal of Service Research*, 3 (3), S. 196-204.
- e/l/s – Institut für Qualitätsentwicklung GmbH (2004): Projekt Remscheid: Effektivität von HZE. Verfügbar unter: <http://www.els-institut.de/downloads.php> [27.05.2009].
- Esch, Franz-Rudolf/Herrmann, Andreas/Sattler, Henrik (2006): *Marketing. Eine managementorientierte Einführung*. München: Vahlen.
- Freise, Dominique C. (2003): *Teilnahme und Methodik bei Patientenbefragungen*. St. Augustin: Asgard-Verlag.
- Greving, Bert (2007): Messen und Skalieren von Sachverhalten. In: Albers, Sönke/Klapper, Daniel/Konradt, Udo/Walter, Achim/Wolf, Joachim (Hg.): *Methodik der empirischen Forschung*. München: Gabler, S. 65-78.
- Hermann, Steffen P. (2006): Stakeholder Management – Long term business success through sustainable stakeholder relationships. In: „Excellence One“, die Learning- und Wissensplattform der EFQM (European Foundation for Quality Management), Mai 2006. Verfügbar unter: http://www.tns-infratest.com/presse/pdf/autorenbeitraege/Hermann_Stakeholder_Management.pdf [07.02.2011].
- Hermann, Steffen P. (2008): Stakeholder based measuring and management of CSR and its impact on corporate reputation. In: Huber, Margit/O’Gorman, Susanne (Hg.): *From customer retention to a holistic stakeholder management system*. Berlin: Springer, S. 51-62.
- Homburg, Christian/Giering, Annette (1996): Konzeptualisierung und Operationalisierung komplexer Konstrukte. In: *Marketing ZFP*, 18 (1), S. 5-24.
- Homburg, Christian/Werner, Harald (1998): *Kundenorientierung mit System. Mit Customer Orientation Management zu profitablen Wachstum*. Frankfurt/M. u.a.: Campus Fachbuch.
- Huber, Margit/Scharioth, Joachim (2003): From customer satisfaction via stakeholder management to the Balanced Scorecard. In: Huber, Margit/Scharioth, Joachim (Hg.): *Achieving excellence in stakeholder management*. Berlin: Springer, S. 1-7.
- Huber, Margit (2008): The emerging market of stakeholder management. In: Huber, Margit/O’Gorman, Susanne (Hg.): *From customer retention to a holistic stakeholder management system*. Berlin: Springer, S. VII-XIII.
- Jarvis, Ian (2008): The beginnings of TRI*M – a personal account. In: Huber, Margit/O’Gorman, Susanne (Hg.): *From Customer Retention to a Holistic Stakeholder Management System*. Berlin: Springer, S. 1-8.
- Koch, Günther/Fertsch-Röver, Jörg (2009): Multiperspektivität als methodische Antwort auf die Komplexität Sozialer Arbeit? In: Maykus, Stephan (Hg.): *Praxisforschung in der Kinder- und Jugendhilfe: Theorie, Praxisbeispiele und Entwicklungsoptionen eines Forschungsfeldes*. Wiesbaden: VS Verlag, S. 123-135.
- Kortendieck, Georg (2008): *Strategisches Controlling in sozialen Organisationen: Studienbrief*. Hochschulverbund Distance Learning.
- Kromrey, Helmut (2007): Begleitforschung und Evaluation – fast das Gleiche, aber doch etwas Anderes! In: Glaser, Michaela/Schuster, Silke (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle: Deutsches Jugendinstitut e.V., S. 113-136.

- Lüders, Christian/Haubrich, Karin (2002): Qualitative Evaluationsforschung. In: Schweppe, Cornelia (Hg.): Qualitative Forschung in der Sozialpädagogik. Wiesbaden: VS Verlag, S. 305-330.
- Martilla, John A./James, John C. (1977): Importance-Performance Analysis. In: *Journal of Marketing*, 41 (1), S. 77-79.
- Matz, Stefanie (2007): Erfolgsfaktorenforschung im Innovationsmanagement. In: Matz, Stefanie (Hg.): Erfolgsfaktoren im Innovationsmanagement von Industriebetrieben. Wiesbaden: Deutscher Universitäts-Verlag, S. 11-49.
- Matzler, Kurt/Sauerwein, Elmar (2002): The factor structure of customer satisfaction: An empirical test of the importance grid and the penalty-reward-contrast analysis. In: *International Journal of Service Industry Management*, 13 (4), S. 314-332.
- Matzler, Kurt/Sauerwein, Elmar/Heischmidt, Kenneth A. (2003): Importance-Performance Analysis revisited: The role of the factor structure of customer satisfaction. In: *The Service Industries Journal*, 23 (2), S. 112-129.
- Matzler, Kurt/Sauerwein, Elmar/Stark, Christian (2008): Methoden zur Identifikation von Basis-, Leistungs- und Begeisterungsfaktoren. In: Hinterhuber, Hans H./Matzler, Kurt (Hg.): Kundenorientierte Unternehmensführung. Wiesbaden: Gabler, S. 321-343.
- Menold, Natalja (2007): Methodische und methodologische Aspekte der Wirkungsmessung. In: Sommerfeld, Peter/Hüttemann, Matthias (Hg.). Hohengehren: Schneider Verlag, S. 26-39.
- Meyer, Wolfgang (2007a): Messen: Indikatoren – Skalen – Indizes – Interpretationen. In: Stockmann, Reinhard (Hg.): Handbuch zur Evaluation: Eine praktische Handlungsanleitung. Münster u.a.: Waxmann.
- Meyer, Wolfgang (2007b): Evaluationsdesigns. In: Stockmann, Reinhard (Hg.): Handbuch zur Evaluation: Eine praktische Handlungsanleitung. Münster u.a.: Waxmann, S. 143-163.
- Moisl, Dominique (2010): Aktivierender Sozialstaat und Herausforderungen für die Evaluationsforschung. In: *Standpunkt Sozial* 2010 (3): S. 7-15.
- Müller, Heinz (2009): Praxisforschung zwischen Erkenntnisgewinn und praktischer Nützlichkeit: Transfer und Transformation als integrale Bestandteile einer „widersprüchlichen Einheit“. In: Maykus, Stephan (Hg.): Praxisforschung in der Kinder- und Jugendhilfe: Theorie, Praxisbeispiele und Entwicklungsoptionen eines Forschungsfeldes. Wiesbaden: VS Verlag, S. 137-151.
- Nagy, Mark S. (2002): Using a single-item approach to measure facet job satisfaction. In: *Journal of Occupational and Organizational Psychology*, 75, S. 77-86.
- Netemeyer, Richard G./Bearden, William O./Sharma, Subhash (2003): *Scaling procedures: issues and applications*. Thousand Oaks: Sage Publications.
- Otto, Hans-Uwe (2007): What works? Expertise. Zum aktuellen Diskurs um Ergebnisse und Wirkungen im Feld der Sozialpädagogik und Sozialarbeit - Literaturvergleich nationaler und internationaler Diskussion. Berlin: AGJ (Arbeitsgemeinschaft für Kinder- und Jugendhilfe).
- Parasuraman, A./Zeithaml, Valarie/Berry, Leonard (1988): SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perception of Service Quality. In: *Journal of Retailing*, 64 (1), S. 12-40.
- Peter, Paul J. (1979): Reliability: A review of psychometric basics and recent marketing practices. In: *Journal of Marketing Research*, 16, S. 6-17.
- Pracht, Arnold/Bachert, Robert (2005): *Strategisches Controlling. Controlling und Rechnungswesen in sozialen Unternehmen*. Weinheim: Juventa.
- Reichheld, Frederick F. (2003): The one number you need to grow. In: *Harvard Business Review*, 81 (12), S. 1-11.
- Rost, Jürgen (2000): Allgemeine Standards für die Evaluationsforschung. In: Hager, Willi/Patry, Jean-Luc/Brezing, Hermann (Hg.): *Evaluation psychologischer Interventionsmaßnahmen: Standards und Kriterien: Ein Handbuch*. Bern: Huber, S. 129-140.
- Sampson, Scott E./Showalter, Michael J. (1999): The performance-importance response function: observations and implications. In: *The Service Industries Journal*, 19 (3), S. 1-25.
- Scharioth, Joachim/Huber, Margit (2003): Introduction. From customer satisfaction via stakeholder management to the Balanced Scorecard. In: Scharioth, Joachim/Huber, Margit (Hg.): *Achieving excellence in stakeholder management*. Berlin u.a.: Springer, S. 1-7.
- Scharioth, Joachim/Huber, Margit/Jarvis, Ian (2004): The Balanced Scorecard as an internal auditing tool. In: Huber, Margit/Scharioth, Joachim/Pallas, Martina (Hg.): *Putting Stakeholder Management into practice*. Berlin u.a.: Springer, S. 135-157.
- Schiepek, Günter/Bauer, Petra (1998): Produktion und Beurteilung von „Qualität“ in psychosozialen Einrichtungen. In: Brunner, Ewald J./Bauer, Petra/Volkmar, Susanne (Hg.): *Soziale Einrichtungen bewerten. Theorie und Praxis der Qualitätssicherung*. Freiburg i.Br.: Lambertus-Verlag.

- Schmidt, Maxie (2008): *Zufriedenheitsorientierte Steuerung des Customer Care*. Wiesbaden: Deutscher Universitäts-Verlag.
- Schnell, Rainer/Hill, Paul B./Esser, Elke (1999): *Methoden der empirischen Sozialforschung*. München: Oldenbourg.
- Schrapper, Christian (2004): *Wie wirkt Sozialpädagogik? Theoretische und methodische Positionen sozialpädagogischer Forschung*. In: Schrapper, Christian (Hg.): *Sozialpädagogische Forschungspraxis: Positionen, Projekte, Perspektiven*. Weinheim: Juventa, S. 11-25.
- Schuster, Silke (2007): *Improvisation, Partizipation und die Frage der Wirkungen – Eine Untersuchung zu Evaluationen der pädagogischen Arbeit gegen Rechtsextremismus*. In: Glaser, Michaela/Schuster, Silke (Hg.): *Evaluation präventiver Praxis gegen Rechtsextremismus. Positionen, Konzepte und Erfahrungen*. Halle: Deutsches Jugendinstitut e.V., S. 169-181.
- Slack, Nigel (1994): *The Importance-Performance Matrix as a determinant of improvement priority*. In: *International Journal of Operations and Production Management*, 14 (5), S. 59-75.
- Sondervan, Eric (2008): *The TRI*M principle – applying it in the public sector*. In: Huber, Margit/O’Gorman, Susanne (Hg.): *From customer retention to a holistic stakeholder management system*. Berlin: Springer, S. 31-50.
- Stauss, Bernd (1999): *Kundenzufriedenheit*. In: *Marketing ZFP*, 21 (1), S. 5-24.
- Stockbauer, Uta (2000a): *Was macht Evaluationen nützlich? Grundlagen und empirische Untersuchungen zum Thema Verwertung und Verwertbarkeit von Evaluationen*. Universität Salzburg: unveröffentlichte Dissertation.
- Stockbauer, Uta (2000b): *Was macht Evaluationen nützlich? Überblick zum Forschungsstand – Ergebnisse von Fallstudien*. In: Müller-Kohlenberg, Hildegard/Münstermann, Klaus (Hg.): *Qualität von Humandienstleistungen. Evaluation und Qualitätsmanagement in Sozialer Arbeit und Gesundheitswesen*. Opladen: Leske + Budrich, S. 121-128.
- Stockmann, Reinhard (2007a): *Einführung in die Evaluation*. In: *Handbuch zur Evaluation*. Münster: Waxmann, S. 24-70.
- Stockmann, Reinhard (2007b): *Evaluation in der Gesellschaft: Entwicklung, Stand und Perspektiven*. In: *Zeitschrift für Evaluation Bd. 2/2007*, S. 195-222.
- Thayer, Colette E./Fine, Allison H. (2001): *Evaluation and outcome measurement in the non-profit sector: stakeholder participation*. In: *Evaluation and Program Planning*, 24 (1), S. 103-108.
- Töpfer, Armin (1996): *Zehn Schritte zur Messung und Steigerung der Kundenzufriedenheit*. In: Töpfer, Armin (Hg.): *Kundenzufriedenheit messen und steigern*. Neuwied u.a.: Luchterhand, S. 229-274.
- Töpfer, Armin/Mann, Andreas (1996): *Kundenzufriedenheit als Meßplatte für den Erfolg*. In: Töpfer, Armin (Hg.): *Kundenzufriedenheit messen und steigern*. Neuwied u.a.: Luchterhand, S. 25-57.
- Töpfer, Armin (2008a): *Konzeptionelle Grundlagen und Messkonzepte für den Kundenzufriedenheitsindex (KZI/CSI) und den Kundenbindungsindex (KBI/CRI)*. In: Töpfer, Armin (Hg.): *Handbuch Kundenmanagement*. Berlin: Springer, S. 309-382.
- Töpfer, Armin (2008b): *Messung von Kunden-Feedback – Ein 10-Punkte-Programm*. In: Töpfer, Armin (Hg.): *Handbuch Kundenmanagement: Anforderungen, Prozesse, Zufriedenheit, Bindung und Wert von Kunden*. Berlin: Springer, S. 419-438.
- Tornow, Harald (2006): *Zufriedenheit von jungen Menschen und ihren Eltern mit stationärer Erziehungshilfe*. In: *EREV Schriftenreihe*, 2006 (3), S. 7-22.
- Vaughan, Liz/Shiu, Edward (2001): *ARCHSECRET: A multi-item scale to measure service quality within the voluntary sector*. In: *International Journal of Nonprofit & Voluntary Sector Marketing*, 6 (2), S. 131-144.
- Wanous, John P./Reichers, Arnon E./Hudy, Michael J. (1997): *Overall job satisfaction: how good are single-item measures?* In: *Journal of Applied Psychology*, 82 (2), S. 247-252.
- Widmer, Thomas (2006): *Qualität in der Evaluation – Wenn Wissenschaft zur praktischen Kunst wird*. In: Stockmann, Reinhard (Hg.): *Evaluationsforschung: Grundlagen und ausgewählte Forschungsfelder*. Münster: Waxmann, S. 15-46.
- Wolff, Stephan/Scheffer, Thomas (2002): *Begleitende Evaluation in sozialen Einrichtungen*. In: Schewpe, Cornelia (Hg.): *Qualitative Forschung in der Sozialpädagogik*. Wiesbaden: VS Verlag, S. 331-351.

8. Anhang

Im Folgenden sind die insgesamt 27 Items aufgeführt, die in die Berechnung der Handlungsrelevanzmatrix eingeflossen sind. Jedes Item wurde gesondert einerseits in Bezug auf Ärzte/Ärztinnen (Ä) und andererseits in Bezug auf Pflegekräfte (P) erhoben. Als Antwortvorgabe diente eine 4-stufige Zustimmungsskala: stimme überhaupt nicht zu (1) – stimme eher nicht zu (2) – stimme eher zu (3) – stimme voll und ganz zu (4). Negativ formulierte Items wurden für die Analyse umgepolt.

Itemliste:

- Die (Ä/P) waren offen und ehrlich zu mir
- Ich hatte vollstes Vertrauen (zu meinen behandelnden Ä/den P) auf der Station
- Die (Ä/P) ließen mich ausreden
- Ich hatte einen festen Ansprechpartner unter den (Ä/P)
- Den (Ä/P) kann man auch Persönliches anvertrauen
- Ich hatte den Eindruck, dass die (Ä/P) fachlich sehr kompetent sind
- Die (Ä/P) ließen mich mit unbeantworteten Fragen zurück
- Die (Ä/P) informierten nur auf Anfrage
- Die (Ä/P) vergaßen schon einmal ein Anliegen, das man hatte
- Die (Ä/P) hatten zu wenig Zeit für mich
- Ich konnte mich auf die (Ä/P) verlassen, wenn ich Probleme mit meiner Krankheit hatte
- Die (Ä/P) unterstützten mich so, dass mir der Umgang mit meiner Krankheit leichter fiel
- Die (Ä/P) waren bereit, sich meine Probleme im Zusammenhang mit der Krankheit anzuhören
- Die (Ä/P) führten auch außerhalb der Visite regelmäßig Gespräche mit mir
- Die (Ä/P) führten die Gespräche mit mir sehr gefühlvoll durch
- Die (Ä/P) ließen mir Zeit, über wichtige Entscheidungen nachzudenken
- Die (Ä/P) waren immer erreichbar
- Die (Ä/P) haben zu oft gewechselt
- Ich hatte den Eindruck, dass es Schwierigkeiten bei der Absprache zwischen den (Ä/P) gab
- Bei den (Ä/P) in diesem Krankenhaus ist man in guten Händen
- Ich wurde von den (Ä/P) auf unpersönliche Weise behandelt
- Ich wurde von den (Ä/P) wie „Luft“ behandelt
- Die (Ä/P) behandelten mich ungerecht
- Ich hatte das Gefühl, dass die (Ä/P) mich persönlich ablehnten
- Es gab Spannungen zwischen mir und den (Ä/P)
- Die (Ä/P) wirkten überheblich auf mich
- Zwischen den (Ä/P) herrschte eine schlechte Stimmung