

# Standards: Aktuelle Strategie zur Qualitätsentwicklung in der Evaluation

Von Wolfgang Beywl und Sandy Taut\*

## Zusammenfassung

*Evaluation wird in der EU und auch in Deutschland verstärkt nachgefragt. Sie tritt eigenständig neben New Public Management und Qualitätsmanagement. Mit Wurzeln bei personenbezogenen Dienstleistungen ist Evaluation über zahlreiche Politikfelder und in zahlreichen organisationalen Kontexten besonders leistungsstark, wenn erwünschte Wirkungen nur dann in befriedigendem Maße erreicht werden, wenn die Bürger hieran aktiv mitwirken. Am Beispiel der aus der Bildungspolitik stammenden amerikanischen ‚Standards der Programmevaluation‘ (JC-Standards) wird veranschaulicht, wie die Qualität von Evaluationen umfassend gesichert und entwickelt werden kann. „Programm“ als generischer Begriff der Evaluationsfachsprache wird geklärt, sowie Inhalt und Aufbau der JC-Standards erläutert. Die Autoren liefern Belege für ihre Anwendbarkeit in der Breite der Politikfelder und stellen Anforderungen an ihre interkulturelle Übertragung vor. Nach der Klärung des Stellenwertes der JC-Standards werden Möglichkeiten ihrer Integration in nationale und internationale Regelwerke zur Evaluation diskutiert. Am Schluss steht das Plädoyer, den Standard-Setzungs-Prozess in Deutschland und Europa schnell voranzubringen.*

## 1. Aktualität von Evaluation und Notwendigkeit von Standards

Erst kürzlich hat Evaluation als wissenschaftlich basierte Vorgehensweise zur systematischen Beschreibung und Bewertung von Politiken (*policies*), Programmen, Projekten, Materialien und Medien auch in den westeuropäischen Staaten Fuß gefasst. Dies gilt zunächst für die Verwaltungsebene der Europäischen Union. Die Europäische Kommission hat im Mai 1996 ein Konzept systematischer Evaluation aller Förderprogramme und Aktionen beschlossen, das durch eine Evaluationsabteilung bei der Finanzkommission koordiniert wird, regelmäßig Evaluationsberichte und -programme herausgibt.<sup>1</sup> Auch in Deutschland lässt sich eine verstärkte Nachfrage nach „Evaluation“ feststellen. Seit der Regierungsübernahme durch die Rot-Grüne Koalition in Berlin werden größere Bundesmodellprogramme regelmäßig durch „Evaluationen“ begleitet.<sup>2</sup> Neuere experimentelle Gesetzgebungen, etwa die zum § 101 a BSHG (Pauschalierung der Sozialhilfe), sind verbunden mit der Auflage, dass sich die Modellversuchsträger an der „Evaluation“ beteiligen.<sup>3</sup>

Manche mögen korrigierend einwenden: Evaluation hat wieder Fuß gefasst. Schließlich standen Ansätze wissenschaftlicher Begleitung oder sozialwissenschaftlicher Po-

litikberatung bereits Ende der 60er bis Mitte der 70er Jahre namentlich in Deutschland, aber auch in vielen anderen westeuropäischen Staaten hoch im Kurs.<sup>4</sup> Auf die da-

\* UNIVATION e.V., Institut für Evaluation und wissenschaftliche Weiterbildung, Zülpicher Str. 58, 50674 Köln. Email: [wolfgang.beywl@uni-koeln.de](mailto:wolfgang.beywl@uni-koeln.de) und [sandy.taut@uni-koeln.de](mailto:sandy.taut@uni-koeln.de)

<sup>1</sup> ‚Evaluationsprogramm‘ meint einen konzeptionellen Rahmen für künftige, mittelfristig durchzuführende Evaluationen; siehe: <http://europa.eu.int/comm/budget/evaluation/en/team.html> [September 2000].

<sup>2</sup> Es wird hier keine Aussage darüber gemacht, dass es neuerdings gerade auf Bundesebene absolut mehr Aktivitäten systematischer Beschreibungen und Bewertungen mittels empirischer Daten gibt; hierüber fehlen Zeitreihenuntersuchungen. Wohl werden solche Aktivitäten neuerdings häufiger mit dem Terminus *technicus* „Evaluation“ bezeichnet, was wiederum eine methodologische Reifung des wissenschaftlichen Ansatzes der Evaluation unterstützt.

<sup>3</sup> Einige weitere Indizien für die neuerliche Ausbreitung von Evaluation in Deutschland: Fast alle Landes-Hochschulgesetze und etwa die Hälfte der Schulgesetze verpflichten inzwischen auf oder empfehlen Evaluationen von Unterricht und Lehre bzw. Forschung. 1997 wurde die Deutsche Gesellschaft für Evaluation gegründet; in der 2000er Koalitionsvereinbarung von SPD und Bündnis 90 / Die Grünen in NRW werden Evaluationen für sechs größere Vorhaben in vier Ressorts angekündigt; einen systematischen Überblick über laufende Evaluation gibt es nicht.

<sup>4</sup> Beeindruckender Ausweis sind die weit über hundert Bände der „Schriften der Kommission für Wirtschaftlichen und Sozialen

malige ‚Planungseuphorie‘ folgte eine Distanzierung von Konzepten der rationalen Planbarkeit und Steuerbarkeit der Politik.<sup>5</sup> Erst zu Beginn der 90er Jahre näherten sich Wissenschaft und Politik mit den Zugangsweisen des betriebswirtschaftlich inspirierten, zentral auf Verfahren des Controlling basierenden „New Public Management“ (NPM) einerseits,<sup>6</sup> des ingenieurwissenschaftlich geprägten „Qualitätsmanagements“<sup>7</sup> andererseits, bei Fragen von Steuerung und Verwaltung, Organisationen und Leistungsprozessen wieder an.

NPM / Controlling, Qualitätsmanagement und Evaluation sind komplementäre Methoden moderner Politik, Verwaltung oder Organisation. Während NPM / Controlling wesentlich monetäre Ressourcen und Ströme im Organisations- und Verwaltungshandeln abbildet und damit Budgets und Leistungen besser planbar macht, weist das Qualitätsmanagement den Weg zu einer effektiven und effizienten ‚kundenorientierten‘ Aufbau- und Ablauforganisation der öffentlichen Verwaltung oder privater Organisationen. Beide zusammen schaffen zentrale *Voraussetzungen* dafür, dass öffentliche Politik, dass privat oder freigemeinnützige Leistungserbringung *wirksam* werden, d.h. dass sie der Bürgerschaft insgesamt oder bestimmten Gruppen Vorteile verschaffen oder darauf hinwirken, bei den Zielgruppen erwünschte Lernprozesse oder Handlungsweisen auszulösen.<sup>8</sup>

Das Alleinstellungsmerkmal der Evaluation besteht darin, dass sie über wissenschaftliche Verfahren verfügt, die Wirkungen und Ergebnisse von Programmen bei den Zielgruppen oder in sozialen Systemen methodisch abgesichert zu beschreiben und zu bewerten sowie Grundlagen für wirkungsorientierte Programmsteuerung zu schaffen.<sup>9</sup>

In vielen Leistungs- und Politikfeldern kommt die intendierte „Wirkung“ oder das gewünschte Ergebnis nur dann in gewünschter Quantität und Qualität zu Stande, wenn die Zielgruppen der Programme hieran aktiv mitwirken. Offensichtlich gilt dies z. B. für Bildung und Erziehung, Kriminal- und Gesundheitsvorsorge bzw. Rehabilitation, Integration in das Einwanderungsland oder in das Erwerbsleben und viele andere Humandienstleistungen.<sup>10</sup> Auch Wirtschaftsförderung, Arbeitsmarktpolitik, nicht wenige umwelt- und energiepolitische Programme werden oft erst wirksam, wenn der Adressat der Programme aktiv wird: Wenn der Einzelne tatsächlich eine selbständige Existenz gründet, wenn er einen Arbeitsplatz in einer anderen Region annimmt, sich zur Anschaffung einer Biotonne entschließt oder den Energieverbrauch in seinem Haushalt senkt. Selbst im hochkomplexen Feld der Forschungs- und Innovationspolitik unterbleibt befriedigender Erfolg oft ohne das Mitwirken der einzelnen Forscher und Forscherinnen. Überall da, wo das Wirksam-Werden von Leistungen in hohem Maße angewiesen ist auf das aktive Mitwirken konkreter Personen, liegt der originäre Beitrag der Evaluation zu

Bewertung, Verbesserung und Steuerung. Betriebswirtschaftliche oder ingenieurwissenschaftliche Ansätze beschränken sich hingegen bei der Messung von Wirkungen bei Zielgruppen oder in sozialen Systemen oft auf grob abschätzende Kennzahlen.<sup>11</sup>

Evaluation findet vielfach Anwendung im Feld personenbezogener Dienstleistungen, seien es Schule und Unterricht,<sup>12</sup> soziale Arbeit,<sup>13</sup> Gesundheitswesen oder Personalentwicklung.<sup>14</sup> Kennzeichnend für Humandienstleistungen ist, dass sowohl in der Fachwelt als auch bei den Beteiligten und Betroffenen in höchstem Maße umstritten ist, welches die Merkmale sind, an denen hohe Qualität von Leistungen gemessen werden soll und welche Ausprägungen dieser Merkmale befriedigende Qualität indizieren. Die nie beendeten Streitigkeiten um integrative Gesamtschulen, Koedukation, Notengebung oder den Numerus Clausus sind beredtes Zeugnis der starken Werteabhängigkeit z. B. von Bildungsevaluationen.

---

Wandel“, die zwischen 1973 und 1996 im Verlag O. Schwartz veröffentlicht wurden.

<sup>5</sup> Vgl. Wollmann (1999).

<sup>6</sup> Vgl. die Beiträge in Blanke et al. (Hrsg.) (1998).

<sup>7</sup> Qualitätsmanagement (QM) in der öffentlichen Politik und Verwaltung ist nicht so gut dokumentiert wie New Public Management / Neues Steuerungsmodell (NSM). QM im öffentlichen Sektor befindet sich im Jahr 2000 auch noch nicht auf dem Höhepunkt seiner Entwicklung. Verwiesen sei auf den Bericht Nr. 7/1995 „Qualitätsmanagement“ der Kommunalen Gemeinschaftsstelle für Verwaltungsvereinfachung (KGSt) Köln und auf Heinz (2000), der mit dem Ansatz „Kommunales Management“ eine Amalgamierung von NSM und QM versucht. Das Land Nordrhein-Westfalen orientiert sich mit seiner durch das Innenministerium koordinierten Initiative „Verwaltungsmodernisierung“ stark an der Systematik der European Foundation for Quality Management (EFQM); vgl. <http://www.moderne-verwaltung.nrw.de/> [September 2000].

<sup>8</sup> Beispiele sind Programme zur Integration straffälliger Jugendlicher, zur Förderung exzellenter Hochschullehre oder zur Steigerung energieeffizienten Wirtschaftens im Unternehmen oder Privathaushalt.

<sup>9</sup> Zu einer ähnlichen Einschätzung kommt Nullmeier (1998), 322.

<sup>10</sup> Die „Humandienstleistungen“ haben über die allgemeinen Charakteristika von Dienstleistungen hinaus (z.B. keine Lagerfähigkeit, uno-actu-Prinzip; vgl. Haller 1998, 52ff) die folgende Besonderheit: Ihre Qualität wird bestimmt durch die je konkret handelnden Personen auf dem Hintergrund von Werten, welche diese Menschen als Individuen innehaben, oder auch als Mitglieder einer sozialen Einrichtung oder des Trägers dieser Einrichtung. Diese Wertebestimmtheit gilt im doppelten Sinne: (1) Die Beteiligten legen die Qualitätsmaßstäbe maßgeblich fest (Soll-Qualität). (2) Die interessengeleiteten Handlungen aller Beteiligten, der Leistungserbringer wie der Leistungsempfänger, sind ausschlaggebender Faktor für Erreichung von Qualität (Ist-Qualität).

<sup>11</sup> Dies gilt in besonderem Maße für das im Anschluss an Kaplan, Norton (1997) aktuell stark in die Diskussion gebrachte Konzept der „Balanced Scorecard“.

<sup>12</sup> Im Reader von Wulf (1972) sind einige Wurzeln nachzulesen; vgl. auch Beywl (1988).

<sup>13</sup> Vgl. Grohmann (1997).

<sup>14</sup> Vgl. Hinweise bei Rossi, Freeman, Lipsey (1999), 10ff.

Systematische Bewertungen und Beschreibungen mit Anspruch auf intersubjektive Nachvollziehbarkeit<sup>15</sup> sind damit hoch verwundbar durch interessengeleiteten Widerspruch. Dieser kann sich auf alle Phasen des Evaluationsprozesses beziehen: Wurden im Evaluationsvertrag einseitig die Interessen des Geldgebers berücksichtigt? Wurden Personen mit der Evaluation beauftragt, die auf Grund ihrer ethnischen Herkunft, ihres Geschlechts oder ihrer politischen Orientierung parteiisch sind? Wurde der Untersuchungsrahmen so gewählt, dass bestimmte unangenehme Befunde von vornherein ausgeblendet sind? Wurden Erhebungsinstrumente eingesetzt, die einen Teil der Zielgruppen (z. B. mit anderer als deutscher Muttersprache) diskriminieren? Wurde bei der Interpretation der Daten einseitig die Werteposition der mächtigen Beteiligten (oder der opponierenden) zu Grunde gelegt? Ist der Bericht so geschrieben, dass die von seinen Schlussfolgerungen Betroffenen ihn nicht verstehen können?

Evaluationen sollen, während sie durchgeführt werden und nachdem sie abgeschlossen sind, auf Akzeptanz und Glaubwürdigkeit bei den am evaluierten Programm Beteiligten und den Adressaten der Evaluationsberichte stoßen. Dies muss gezielt vorbereitet und den Adressaten kommuniziert sein. Diejenigen, die bei Gelingen oder Misslingen von Evaluationen kurzfristig am meisten zu verlieren haben, sind zum einen die Auftraggeber bzw. Geldgeber und zum anderen die Evaluatoren und Evaluatortinnen. Längerfristig sind auch die Zielgruppen von Programmen oder die Steuerzahler davon tangiert, ob Evaluationen gut durchgeführt werden, ob sie beabsichtigte Verbesserungen von Programmen unterstützen bzw. zu angemessenen Entscheidungen führen.

Ein im Fortgang dieses Beitrags ausgeführter Weg, die Qualität und damit auch Akzeptanz und Nutzung von Evaluation zu erhöhen, besteht darin, Kriterien, Regeln, Leitlinien oder Standards zu formulieren, denen gute Evaluationen genügen sollen. Weitere Qualitätsaspekte, z. B. die Ausbildung des Evaluationspersonals oder die Anforderungen an Evaluationsaufträge betreffend, können durch solche Regelwerke<sup>16</sup> mittelbar gefördert werden, indem sie z. B. in Ausbildungsordnungen an Hochschulen oder Richtlinien für öffentliche Ausschreibungen aufgenommen werden.

In den zurückliegenden zwei Jahrzehnten sind — zunächst in den USA — verschiedene Regelwerke zur Evaluation entwickelt und verabschiedet worden. Grob unterscheiden lassen sich solche, die auf das Handeln der Evaluationsverantwortlichen (Evaluatoren, aber auch Auftraggeber) abzielen und solche, die Anforderungen an Evaluationen als fachliche Leistungsprozesse formulieren.<sup>17</sup> Im Folgenden sollen als prominentes Beispiel die Standards der Programmevaluation des *Joint Committee on Standards for Educational Evaluation* vorgestellt werden. Dabei wird sowohl ihre Übertragbarkeit auf Politikfelder über Bildung / Erziehung hinaus sowie auf andere kul-

turelle Kontexte, als auch ihre Fortentwicklung zu und Passung mit ersten europäischen Evaluationsstandards diskutiert.

## 2. Die Standards der Programmevaluation (JC-Standards) und die Möglichkeiten ihrer Anwendung und Übertragung

In den USA gründete sich 1975 ein interdisziplinäres Komitee, um Standards für die Evaluation in Bildung und Erziehung zu entwickeln, das *Joint Committee on Standards for Educational Evaluation*. Die zwölf beteiligten Organisationen veröffentlichten 1981 die *Standards for Evaluations of Educational Programs, Projects, and Materials* (publiziert bei McGraw-Hill). Das Regelwerk wurde 1989 als „Nationaler Standard“ vom *American National Standards Institute* (ANSI) anerkannt. Im Jahre 1988 folgte die Publikation der *Personnel Evaluation Standards* (bei Sage), die für die Personalevaluation insbesondere an Schulen und Hochschulen einen Rahmen bieten. 1994 ist nach umfassender Überarbeitung und fachöffentlicher Überprüfung eine von nun 16 Organisationen<sup>18</sup> getragene Neuauflage der 81er Standards erschienen. Die neue Publikation führt den Titel *Program Evaluation Standards* und deutet den Anspruch der Joint-Committee-Standards (nachfolgend kurz: JC-Standards) an, auch über den Bereich von Bildung und Erziehung hinaus zu gelten.<sup>19</sup> Im Titel der 1999 erstmalig erschienenen deutschen Übersetzung *Handbuch der Evaluationsstandards* ist der Bezug zu einem speziellen Anwendungsbereich aufgelöst.

### 2.1 Zum Begriff „Programm“ in der Evaluation

Eine erste Klippe bei der Verbreitung der JC-Standards ist die Übertragung des Begriffes *program* in den deutschen Kontext. Je nach Disziplin oder Politikfeld hat „Programm“ in Deutschland eine andere Bedeutung.<sup>20</sup> Auf der

<sup>15</sup> In der Alltagssprache oder auch in Ausschreibungen zu Evaluationen wird hier häufig die Bezeichnung „Objektivität“ genutzt. Evaluation kann nie objektiv im Sinne von ‚über den Wertkonflikten oder außerhalb von ihnen stehend‘ sein. Sie kann Werte transparent und bewusst machen.

<sup>16</sup> Der Terminus „Regelwerke“ wird hier zur zusammenfassenden Bezeichnung von „Standards“, „Leitprinzipien“ usw. verwendet, in denen die Regelungsdichte teils sehr niedrig, teils hoch ist.

<sup>17</sup> Eine Übersicht über insgesamt zehn evaluationsspezifische Regelwerke geben Beywl, Widmer (2000).

<sup>18</sup> Darunter American Association of School Administrators, American Educational Research Association, American Evaluation Association, American Federation of Teachers, Council of Chief State School Officers.

<sup>19</sup> Der Ursprung ist noch im Untertitel zu erkennen: „How to Assess Evaluations of Educational Programs“.

<sup>20</sup> Die folgende Argumentation versucht einen Programm-Begriff zu umreißen, der für Programm-Evaluationen aller Art

Ebene internationaler oder nationaler Politik z. B. bezeichnet „Programm“ größere Bündel von Maßnahmen oder Teilstrategien wie z. B. die insgesamt ca. 30 Milliarden Euro umfassenden Programme der Strukturpolitischen Maßnahmen der Europäischen Union.

In den USA ist *program* ein generischer Begriff der Evaluationsfachsprache. Scriven fasst ein Programm als „*the general effort that marshals staff and projects toward some (often poorly) defined and funded goal*“.<sup>21</sup> Andere Autoren betonen ebenfalls die Zielgerichtetheit eines Maßnahmenbündels, welches im Zusammenspiel von Planung und Durchführung von Aktivitäten und zu erreichenden Ergebnissen ein Programm ausmacht.<sup>22</sup>

Der Komplexität eines Programms werden dabei nach oben Grenzen gesetzt durch den Übergang in eine *policy*. Diese bezeichnet das staatliche und parastaatliche Handeln in einem in Bezug auf Ziele lediglich grob umrissenen Politikfeld, z. B. der Sozial-, Bildungs- oder Gesundheitspolitik. In einem Politikfeld handeln Parlamente, Verwaltungen, Sozialversicherungsträger, Kammern, Innungen, Berufsverbände, Verbraucherzusammenschlüsse usw. arbeitsteilig und z. T. autonom voneinander. In einem Programm hingegen gibt es idealtypisch oberste Entscheider (dies kann auch ein demokratisch legitimiertes Gremium sein), Programmverantwortliche oder -manager, Mitarbeiter und so etwas wie Zielgruppen (Einzelpersonen, Unternehmen, Kommunen). Faktisch werden besonders große „Makro-Programme“ (z. B. die Strukturfondsprogramme der EU oder der Kinder- und Jugendplan des Bundes) zwar zentral budgetiert, aber in einem mehrstufigen Verfahren schließlich regional alloziiert, was ihre zielgeführte Steuerung und auch Evaluation erheblich kompliziert.

Vedung weist darauf hin, dass einem Programm in aller Regel mehrere Programmtheorien zu Grunde liegen, also z. B. die der Entscheider und die der Umsetzer.<sup>23</sup> Fallen die Programmtheorien mit ihren Zielen und Bewertungskriterien stark auseinander, ergeben sich für die Evaluierbarkeit von Programmen ähnliche Schwierigkeiten wie für die von *policies*.<sup>24</sup>

Nach unten hin werden Programme im engeren Sinne öfter gegenüber den zeitlich und in ihrer Reichweite begrenzteren Projekten abgegrenzt.<sup>25</sup> Im Rahmen von größeren Programmen treten beispielsweise Werbe- und PR-Kampagnen, öffentliche Veranstaltungen, Seminare und Kurse, Unterrichtsreihen, eine Folge von Beratungsterminen auf, die alle für sich genommen programmförmig sind. Wir bezeichnen sie daher als Meso-Programme im weiteren Sinne: Auch hier werden auf der Basis von Ressourcen, ausgerichtet auf bestimmte (im Idealfall operationalisierte) Ziele, Aktivitäten entfaltet, die zu (befriedigenden oder nicht befriedigenden) Resultaten führen. Wie die Evaluationen mehrerer lokaler Programme / Projekte, die im Rahmen eines übergreifenden Programms durchge-

führt werden, zusammengeführt werden können, wird neuerdings verstärkt unter dem Stichwort „Cluster-Evaluation“ diskutiert.<sup>26</sup>

Programme kommen schließlich auch auf einem Mikro-Level etwa therapeutischer oder intensivpädagogischer Prozesse zwischen zwei Personen zu Stande. Das *Handbuch der Evaluation psychologischer Interventionsmaßnahmen* nutzt „Programm“ als Synonym für „Maßnahme“ und „Intervention“ im Bereich der klinischen, der pädagogischen und der Organisations-Psychologie.<sup>27</sup> Auch für diese Bereiche unterscheiden die Autoren zwischen „flächendeckenden“ und „lokalen“ Programmen.<sup>28</sup>

Für die Evaluation in den USA (und auch in Australien und Kanada) gilt, dass Gebrauch wie Verstehenskonsens des Begriffes *program* höher sind als in Deutschland. Dies ist vor der Kulisse der Politikmuster in diesen Ländern nicht verwunderlich. Die amerikanische Regierung nähert sich seit den 60er Jahren (im Rahmen von Kennedys „War against Poverty“ und Johnsons „Great Society“) vielen sozialen Problemen über Programme, die spezifische Zielgruppen (z. B. Minoritäten, Bewohner strukturschwacher Gebiete etc.) ansprechen sollen und deren Weiterfinanzierung von ihrem Erfolg in Bezug auf gesetzte Ziele abhängig ist. In Deutschland war die Problembearbeitung hingegen viel längere Zeit institutionalisiert, d. h. die öffentliche Hand bzw. von ihr beauftragte Einrichtungen erbrachten die Leistungen auf Basis einer dauerhaften Finanzierung oft in Art eines Monopols. Ihr Aufgabenbereich war abgesteckt; ausgewiesene Ziele fehlten oder waren wenig präzisiert oder wurden zumindest nicht auf ihre Erreichung überprüft. Mit NPM / Neuer Steuerung, der Erfordernis, Kontrakte (intern und extern) auf der Basis von Leistungs-, Entgelt- und Qualitätsvereinbarungen zu treffen,<sup>29</sup> löst sich institutionalisierte vielfach in pro-

---

anwendbar ist. Er kollidiert vielfach mit disziplinär gebundenen Programm-Begriffen (etwa der Politikwissenschaft oder der Erziehungswissenschaft); die begriffliche Klärung muss vorläufig bleiben und bedarf intensiver interdisziplinärer Diskussion.

<sup>21</sup> Scriven (1991), 285.

<sup>22</sup> Vgl. z. B. Rossi, Freeman, Lipsey (1999); Worthen, Sanders, Fitzpatrick (1997).

<sup>23</sup> Vedung (1999), 124.

<sup>24</sup> Als Vorgehensweise steht die Evaluierbarkeitsabschätzung zur Verfügung, die oft zu dem Ergebnis führen wird, zunächst im Rahmen einer formativen Evaluation einen Klärungsprozess der zu Grunde zu legenden Programmtheorie durchzuführen. Vgl. Patton (1997), 152f.; grundlegend vgl. Wholey, Hatry, Newcomer (1994).

<sup>25</sup> So auch in den JC-Standards (1994), 3: „Program ... activities that are provided on an continuing basis“ vs. „Project ... activities that are provided for a defined period of time ... Projects that become institutionalized become programs.“

<sup>26</sup> Vgl. Sanders (1997).

<sup>27</sup> Hager, Patry, Brezing (2000).

<sup>28</sup> Ebenda, 142.

<sup>29</sup> So z. B. geregelt in Paragraph 78a-f des Kinder- und Jugendhilfegesetzes für die stationäre und teilstationäre Erziehungshilfe.

grammförmige Leistungserbringung auf. Programmevaluation tritt auf den Plan.

Wir schlagen vor, „Programm“ als generischen Begriff der Evaluationsfachsprache zu nutzen, der von der Mikro-Ebene zielgerichteter, z. B. therapeutischer, Interventionsbündel bis zur Makro-Ebene internationaler Programmarchitekturen anwendbar ist und definieren ihn wie folgt:

Programme sind beschriebene und durchgeführte, intentional aufeinander bezogene Bündel von Interventionen, Maßnahmen, Projekten oder Teilprogrammen, die aus einer Folge von auf ausgewiesene Ziele hin ausgerichteten Aktivitäten / Interaktionen bestehen, welche auf der Basis von verfügbaren Ressourcen durchgeführt werden und darauf gerichtet sind, vermittels bereitgestellter Leistungen (outputs) bestimmte, bei bezeichneten Zielgruppen oder im sozialen System zu erreichende Ergebnisse (outcomes) auszulösen. Gegenstand der Evaluation können sowohl das Konzept des Programms (Programm als Plan) als auch die Durchführung des Programms und seine Resultate (Programm in Aktion) sein.

## 2.2 Inhalt und Aufbau der Standards der Programmevaluation

Die JC-Standards sind in Buchform auf ca. 230 Druckseiten veröffentlicht.<sup>30</sup> In der Einführung werden zunächst zentrale Begriffe definiert (u.a. Evaluation, Programm, Evaluationsstandard). Als Zwecke der JC-Standards werden genannt: Hilfestellung bei Planung, Durchführung und kritischer Überprüfung von Evaluationen; Basis für einen orientierten Dialog zwischen den relevanten Beteiligten und Betroffenen über die Qualität von Evaluationen; Information über verschiedenste Methoden insbesondere der Datenerhebung; Positionierung von Evaluationen im politischen Machtgeflecht; schließlich auch Grundlage für die Aus- und Fortbildung in Evaluation. Vorrangige Adressaten der JC-Standards sind die Auftraggeber und die EvaluatorInnen, darüber hinaus Hochschullehrer und Studierende. Auch andere durch Evaluationsergebnisse betroffene Gruppierungen, besonders die Zielgruppen der evaluierten Programme, sind angesprochen.<sup>31</sup> Nach einer kurzen Vorstellung des Gliederungsschemas und des Formats der nachfolgenden 30 Einzelstandards wird auf zehn Seiten die Anwendung der Standards inklusive eines Fallbeispiels (aus der beruflichen Weiterbildung) demonstriert. Die JC-Standards enthalten explizit „... Ratschläge dazu, wie die Angemessenheit von Evaluations-tätigkeiten beurteilt werden kann; sie stellen jedoch keine speziellen Kriterien bereit, anhand derer solche Beurteilungen vorgenommen werden können.“<sup>32</sup>

Den Hauptteil des Buches machen die insgesamt 30 Einzelstandards aus, die in folgende vier Gruppen gegliedert sind:

- Sieben Nützlichkeitsstandards, die sicher stellen sollen, dass sich eine Evaluation an den Informationsbedürfnissen der vorgesehenen Evaluationsnutzer ausrichtet.
- Drei Durchführbarkeitsstandards, die sicher stellen sollen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst ausgeführt wird.
- Acht Korrektheitsstandards, die sicher stellen sollen, dass eine Evaluation rechtlich und ethisch korrekt durchgeführt wird und dem Wohlergehen der in die Evaluation einbezogenen und auch der durch die Ergebnisse betroffenen Personen gebührende Aufmerksamkeit widmet.
- Zwölf Genauigkeitsstandards, die sicher stellen sollen, dass eine Evaluation über die Güte und/oder die Verwendbarkeit des evaluierten Programms fachlich angemessene Informationen hervorbringt und vermittelt.

Damit sind vier zentrale, in der theoretischen Literatur zur Evaluation und zahlreichen Metaevaluationen immer wieder angesprochene Qualitätsdimensionen von Evaluation benannt.<sup>33</sup> Um speziell Evaluatoren und Evaluatorinnen zu unterstützen, Evaluationen Schritt für Schritt mit Hilfe der JC-Standards zu planen, ist ein funktionales Inhaltsverzeichnis vorangestellt, das die Standards zehn Hauptaufgaben einer Programmevaluation zuordnet.<sup>34</sup>

Die dreißig Einzelstandards werden auf je ca. fünf Druckseiten in einer einheitlichen Form präsentiert, um das Lesen zu erleichtern und die Verknüpfung der Standards untereinander zu verdeutlichen:

- *Benennung*: Der beschreibende Titel des Standards, z. B. Kontextanalyse.

<sup>30</sup> Diese Umfangsangabe gilt sowohl für die amerikanische Originalausgabe von 1994 wie für die deutsche Übersetzung von 2000, nach der im Folgenden zitiert wird.

<sup>31</sup> „Die Standards wurden entwickelt, um von Lehrern, Verwaltungspersonal, Mitgliedern der Schulbehörden, Ausbildern, Trainern, Evaluatorinnen, Curriculumexperten, Gesetzgebern, Personalentwicklern, Beratern, Kommunalpolitikern sowie von wirtschaftlichen und pädagogischen Verbänden, Eltern und weiteren Personen verwendet zu werden“ (Joint Committee 2000, 23).

<sup>32</sup> Joint Committee (2000), 26. Die Standards sind damit keine operational definierten Minimalstandards im Sinne des Qualitätsmanagements, die nicht unterschritten werden dürfen. Vielmehr handelt es sich um Standards im sozialwissenschaftlichen Verständnis, also explizite Angaben des Gewünschten (als konkrete Anwendung von Werten / Spezifizierung von Normen). Die terminologische Missverständlichkeit muss u.E. hingenommen werden, da alternative Termini, z. B. „Regel“ oder „Norm“, ebenfalls Mehrdeutigkeiten aufweisen und in den verschiedenen Fachsprachen höchst unterschiedlich konnotiert sind.

<sup>33</sup> Vgl. die Ausführungen und Literaturhinweise zum „Nutzendefizit“ und „Werte-Daten-Dilemma“ in Beywl (1988), 32–41.

<sup>34</sup> U.a. Entscheidung über die Durchführung einer Evaluation, Definition des Evaluationsproblems, Planung der Evaluation, Informationsgewinnung usw. bis Leiten der Evaluation und personelle Ausstattung der Evaluation.

- *Standarddefinition*: Kurzfassung des Standards in Form einer „Soll“-Aussage, z. B.: „Der Kontext, in dem das Programm angesiedelt ist, sollte ausreichend detailliert untersucht werden, damit mögliche Beeinflussungen des Programms identifiziert werden können.“
- *Überblick*: Eine Begriffsklärung/Einführung, bei der Schlüsselbegriffe des Standards definiert und allgemeine Hinweise für seine Verwendung gegeben werden.
- *Richtlinien*: Diese Vorschläge sollen die Evaluatoren und Evaluatorinnen und Beteiligten und Betroffenen orientieren, auf welchen Wegen sie den Erfordernissen des Standards entsprechen können. Sie dienen der Fehlervermeidung und sind weder als erschöpfend noch als zwingend zu beachten.
- *Häufige Fehler*: Eine Liste typischerweise auftauchender Schwierigkeiten wie Fehler, die von unerfahrenen Evaluatoren und Evaluatoren begangen werden oder solche, zu denen es wegen übereifriger Anwendung des Standards kommt.
- *Anschauungsbeispiel*: Ein oder zwei Beispiele dazu, wie der Standard angewendet werden kann. Beschrieben werden das jeweilige Umfeld, eine Situation, bei welcher der Standard (nicht) beachtet wurde, und es erfolgt gegebenenfalls eine Diskussion von Korrekturmaßnahmen, die zur Einhaltung des Standards führen können. Die meisten Anschauungsbeispiele beruhen auf Evaluationen aus der Praxis.

### 2.3 Anwendbarkeit der Standards in verschiedenen Politikfeldern

Wie anfangs beschrieben, sind die JC-Standards ursprünglich für den Bereich Bildung und Erziehung entwickelt worden. Die im Buch abgedruckten ca. 50 Anschauungsbeispiele kommen vorwiegend aus dem Bereich des Elementar- und Sekundarschulwesens, der Colleges und Universitäten, darüber hinaus auch der beruflichen Bildung und der Sozialen Arbeit. Eine Durchsicht der Benennungen und Definitionen der dreißig Einzelstandards ergibt, dass bis auf eine Ausnahme<sup>35</sup> keine spezifischen, auf den Bereich Bildung und Erziehung eingegrenzten Fachwörter benutzt werden. So wird allgemein von Zielgruppen, Adressaten, Mitarbeitern (*staff*) gesprochen. Stockdill untersuchte mittels Experteninterviews, ob die JC-Standards für Evaluationen in der US-amerikanischen Geschäftswelt angemessen sind.<sup>36</sup> Sie kommt zu dem Ergebnis, dass die Standards auch für Personalentwicklung und Evaluation anderer HRD-Aufgaben im Profit-Sektor anwendbar sind. Caracelli und Riggan, die sich mit der Entwicklung von Qualitätskriterien für Evaluationen in den unterschiedlichsten Gegenstandsfeldern befassen, ziehen die JC-Standards bei, ohne auf Probleme bei der Anwendung auf Bereiche außerhalb von Bildung und Erziehung zu stoßen.<sup>37</sup>

Als bahnbrechend für die Frage der intersektoralen Übertragbarkeit der JC-Standards hat sich schließlich die umfassende Meta-Evaluation von Widmer erwiesen.<sup>38</sup> Aus der Grundgesamtheit der insgesamt 53 in den Jahren 1990–1994 abgeschlossenen schweizerischen Evaluationsstudien wurden insgesamt zehn ausführlich beschrieben und mit Hilfe der JC-Standards bewertet. Die Fallstudien betreffen Evaluationen zur Arbeitsschutz-, Umwelt-, Abfall- und Energie- sowie Verkehrspolitik, zur Wohnungsbau-, Wirtschaftsförderungs- und Kulturpolitik. Auf Basis einer Synthese der zehn Meta-Evaluationen werden die zu Grunde gelegten JC-Standards als übertragbar sowohl auf den Schweizer Kontext als auch über den Bildungsbereich hinaus auf andere Politikbereiche eingeschätzt.

Die schon von Widmer vorgeschlagene Straffung der JC-Standards und insbesondere die empfohlene Streichung des Standards K1 werden in den aktuellen Empfehlungen der Schweizerischen Evaluationsgesellschaft (SEVAL) aufgegriffen, die eine begründete Adaption der JC-Standards für die Schweizerische Evaluation darstellen.<sup>39</sup> Die SEVAL stellte auf einer Tagung im Mai 2000 in Bern die im Auftrag der SEVAL ausgearbeiteten Standards vor und diskutierte unter Beteiligung von Auftraggebern und Evaluationsspezialisten aus verschiedenen Politikfeldern, inwiefern diese Standards zu einer Verbesserung der Ausbildung von Evaluatoren und Evaluatorinnen, sowie zur Planung und der Durchführung von Evaluationsprojekten beitragen können. Es wurde sowohl von Auftraggeber- wie auch Auftragnehmerseite in verschiedenen Voten darauf hingewiesen, dass die SEVAL-Standards eine gute Arbeitsgrundlage bilden, um einen konkreten Evaluationsauftrag zu klären und den Evaluationsprozess daran zu orientieren. Schon vor der eigentlichen Verabschiedung der SEVAL-Standards durch die Mitgliederversammlung der SEVAL nutzen einzelne Bundesbehörden dieses Regelwerk, um Evaluationsaufträge zu qualifizieren und Meta-Evaluationen durchzuführen.<sup>40</sup>

Viele Indizien sprechen dafür, dass die Anwendbarkeit der JC-Standards über den Entstehungskontext Bildung

<sup>35</sup> Es ist der Standard K1 „Unterstützung der Dienstleistungsorientierung“, in dem von Evaluationen gefordert wird, sie sollten „unterstützend dafür sorgen, dass Erziehungs- und Bildungsziele angemessen sind“ (Joint Committee 2000, 109). Folgerichtig wurde dieser Standard im Papier der Schweizerischen Evaluationsgesellschaft gestrichen.

<sup>36</sup> Stockdill (1986).

<sup>37</sup> Caracelli, Riggan (1994); vgl. zum Thema der politikfeldübergreifenden Anwendung auch Widmer, Beywl (2000).

<sup>38</sup> Widmer (1996).

<sup>39</sup> Widmer, Landert, Bachmann (1999).

<sup>40</sup> So lässt das Schweizerische Bundesamt für Energie (BFE) die Qualität seiner bisherigen Evaluationsanstrengungen im Rahmen einer international ausgeschriebenen Meta-Evaluation bewerten, die auf den SEVAL-Standards basieren soll.

und Erziehung hinaus auf andere Politikfelder grundsätzlich gegeben ist.

#### 2.4 Interkulturelle Übertragbarkeit

Wie eben gezeigt, werden die US-amerikanischen JC-Standards nicht zuletzt wegen der Kostenintensität eines solchen Standardsetzungsvorganges als Modell für andere nationale Evaluationsstandards diskutiert und verwendet. Die Frage nach der interkulturellen Übertragbarkeit der Standards stellt sich offensichtlich, wird aber schnell vernachlässigt, da im Zeitalter der Globalisierung von Waren-, Geld und Informationsaustausch eine Art dominante ‚Weltkultur‘ als gesetzt gilt, deren Verständigungsmittel die englische Sprache ist.

Standards sind generell stark durch Normen und Werte (der Personengruppe, die für ihre Entstehung verantwortlich ist) geprägt, die von Kultur zu Kultur stark variieren können.<sup>41</sup> Stufflebeam, langjähriger Vorsitzender des Joint Committee, weist auf die Möglichkeit hin, dass die Standards außerhalb der Vereinigten Staaten nur eingeschränkt anwendbar sein könnten.<sup>42</sup> Gleichzeitig unterstützt das Joint Committee die Nutzung der Standards außerhalb der USA.<sup>43</sup> Einige im internationalen Raum tätige Evaluationsexperten halten es ebenfalls für erforderlich, eine (manchmal geringfügige) Anpassung der Standards an die Gegebenheiten des jeweiligen Landes vorzunehmen.<sup>44</sup> In einem von UNICEF organisierten Evaluations-Workshop in Nairobi, Kenia, wurden die Standards auf ihre Anwendbarkeit in Afrika im Rahmen einer Expertendiskussion überprüft und von der Mehrheit der Teilnehmenden nach geringfügigen Anpassungen für übertragbar befunden.<sup>45</sup> Eine Ausgabe der Occasional Paper Series des Evaluation Center (Western Michigan University, USA), das auch das Joint Committee betreut, enthält sechs Beiträge zum Thema *The Program Evaluation Standards in International Settings*.<sup>46</sup> Die Artikel betreffen die Anwendung der Standards in Malta, Indien<sup>47</sup>, Südkorea<sup>48</sup> und der Schweiz<sup>49</sup>.

Eine Prüfung der Wert- und damit Kulturabhängigkeit der dreißig JC-Standards verfolgt Taut im selben Compendium.<sup>50</sup> Basierend auf ethnopsychologischer und anthropologischer Literatur analysiert die Autorin jeden Standard auf der Grundlage empirisch bewährter Wertdimensionen, wie z. B. Kollektivismus-Individualismus, die zur Beschreibung und Kategorisierung von Kulturen verwendet werden.<sup>51</sup> Zusätzlich werden international veröffentlichte Beiträge zum Thema der interkulturellen Übertragbarkeit der Standards berücksichtigt.<sup>52</sup> Die Analyse will zum einen generell für interkulturelle Einflüsse sensibilisieren, zum anderen professionellen Evaluationsgemeinschaften außerhalb der USA Hinweise geben, an welchen Stellen — je nach Kulturkreis — eine wegen Wertunterschieden erforderliche Adaptation der JC-Standards angedacht ist, wenn sie als Modell im Rahmen eines natio-

nen oder regionalen Standardsetzungsprozesses in wiederum kulturell gültige Regelwerke münden sollen.

Beispielsweise wird für den ersten Nützlichkeitsstandard „Identifizierung von Beteiligten“ (*Stakeholder Identification*) die Werte-Dimension Hierarchie-Gleichstellung herangezogen, die ein Kontinuum zwischen Legitimität von hierarchischen Strukturen auf der einen und egalitärem Gesellschaftsverständnis auf der anderen Seite beschreibt.<sup>53</sup> Der Standard fordert dazu auf, alle Beteiligten und Betroffenen eines Evaluationsvorhabens in den Evaluationsprozess einzubinden. Von einem kulturvergleichenden Standpunkt aus betrachtet hat dieser Standard u. a. in seinem US-amerikanischen, eher egalitären Entstehungszusammenhang Berechtigung, weniger allerdings zum Beispiel in Ländern des asiatischen Kulturkreises, wo statische Hierarchien einen unverzichtbaren Bestandteil des Lebens darstellen. Eine Beteiligung von Betroffenen, die im Hierarchiegefüge ‚unten‘ stehen, würde dem herrschenden Kulturverständnis widersprechen und letztlich die Akzeptanz und Glaubwürdigkeit von Evaluation mindern. Demzufolge sollte dieser Standard in solchen Kulturen nicht als unbedingte Zielvorstellung aufrecht erhalten werden.<sup>54</sup>

Die mit den beiden Standardgruppen Nützlichkeit (*utility*) und Korrektheit (*propriety*) verbundenen Sollensaussagen der JC-Standards sind generell stärker wert- und kulturabhängig als Durchführbarkeits (*feasibility*)- und Genauigkeitsstandards (*accuracy*). Bei der nationalen / regionalen Anpassung der Standards sollte die jeweilige Kultur nach den erläuterten Wertdimensionen beschrieben werden, gefolgt von einer systematischen Substitution oder Anpassung derjenigen Standards, die den jeweiligen kulturellen Gegebenheiten nicht oder nur bedingt Rechnung tragen. Je kulturell inhomogener der durch die Standards abzudeckende Bereich ist (z. B. Evaluationsstandards für „Eu-

<sup>41</sup> Berry, Segall, Kagitcibasi (1997), 79f.

<sup>42</sup> Stufflebeam (1986), 18.

<sup>43</sup> Als auf Stufflebeam folgender Vorsitzender des Joint Committee vgl. Sanders (2000), 237.

<sup>44</sup> Z. B. Nevo (1984); Marklund (1984); Smith, Chircop, Mukherjee (1993); Nowacki (2000).

<sup>45</sup> Russon, Patel (1999).

<sup>46</sup> Russon (Hrsg.) (2000).

<sup>47</sup> Smith et al. (1993).

<sup>48</sup> Jang (2000).

<sup>49</sup> Widmer (1995) und ders. (2000).

<sup>50</sup> Taut (2000).

<sup>51</sup> Vgl. Hofstede (1980); Triandis (1995).

<sup>52</sup> Z. B. Smith et al. (1993); Levy (1984); Beywl (1999).

<sup>53</sup> Vgl. Schwartz (1994a).

<sup>54</sup> Ähnlich rät Nowacki (2000), 65ff zu umsichtiger Anwendung der auf Partizipation zielenden Standards in der peruanischen Gesellschaft. Sie hat als Deutsche die Anwendbarkeit der JC-Standards in einem Projekt der Entwicklungszusammenarbeit in Peru überprüft.

ropa“), desto unspezifischer müssen deren Formulierungen ausfallen und desto weniger können sie als konkrete Zielvorgaben für den einzelnen Evaluator wirklich nützlich sein.

Wie eingangs ausgeführt, sind Standards per se wertgebunden und daher kulturspezifisch. Natürlich bewegt man sich hierbei auf einem Kontinuum. Es liegt auf der Hand, dass eine Übertragung der JC-Standards auf z. B. den britischen Kontext weniger Adaptationsarbeiten verlangt als dies von der koreanischen Evaluationsgemeinschaft zu leisten wäre.<sup>55</sup> Für Deutschland darf — ähnlich wie für die Schweiz — davon ausgegangen werden, dass insgesamt geringere Anpassungsnotwendigkeiten bestehen.

Ein spezielles Thema bildet die Verwendung der JC-Standards in der Entwicklungszusammenarbeit, bei der Evaluatoren, Auftraggeber und Beteiligte ein Konglomerat verschiedener Kulturen darstellen. Hier sind die bereits in den JC-Standards vorhandenen Ansätze<sup>56</sup> auszuweiten und Standards so zu formulieren, dass sie besonderes Augenmerk auf kulturelle Sensibilität und Verständigung legen.<sup>57</sup>

### 3. Aktueller Stellenwert der Standards der Programmevaluation

Die faktische Nutzung der JC-Standards in ihrem Ursprungsland ist — in Relation zum Aufwand, der zur Erstellung und Revision getrieben wurde — schlecht dokumentiert. Das Resultat der Suche nach Validierungsstudien der Standards in den USA — außerhalb der Arbeit des Joint Committee, in dessen Rahmen allerdings speziell eingerichtete Validierungsgruppen tätig waren<sup>58</sup> — fällt eher spärlich aus. Auch Berichte über deren tatsächliche Verbreitung bzw. Anwendung in ihrem US-amerikanischen Entstehungskontext sind u. W. nicht existent.<sup>59</sup>

In ihren *Guiding Principles*<sup>60</sup> verweist die *American Evaluation Association* (AEA) explizit auf die JC-Standards, deren Entwicklung und Anwendung sie unterstützt. Zwei stark verbreitete Lehrbücher<sup>61</sup> der Evaluation behandeln die Standards explizit: Rossi, Freeman und Lipsey weisen in der 6. Auflage ihres Lehrbuches positiv auf die JC-Standards hin, bezeichnen sie allerdings als noch nicht voll ausgereift, besonders in Hinblick auf einen geeigneten Rahmen zur Durchsetzung (*enforcement*) der Standards bei Auftraggebern und Evaluatoren.<sup>62</sup> Patton stellt die JC-Standards als zentrale Entwicklungsleistung der Evaluationsdisziplin heraus und baut sie ein in das Fundament seiner *Utilization-Focused Evaluation*.<sup>63</sup> Owen und Rogers stellen die JC-Standards in ihrem neuen Lehrbuch zur Programmevaluation ausführlich vor und empfehlen ihre Anwendung für die Evaluationsplanung und die Evaluationsausbildung ebenso wie für die Meta-Evaluation.<sup>64</sup> Die *Australasian Evaluation Society* (AES)

hat die JC-Standards im August 1996 gutgeheißen. Die AES bezieht sich auch im Zusammenhang mit ihren Ethik-Richtlinien auf die JC-Standards.<sup>65</sup> Es existieren spanische und deutsche Übersetzungen der Langfassung der JC-Standards.<sup>66</sup>

Einschlägige *deutsche* Evaluationslehrbücher erwähnen die JC-Standards entweder gar nicht<sup>67</sup> oder lediglich randständig<sup>68</sup>. Auch das aus dem Schwedischen / Englischen übersetzte Lehrbuch von Vedung erwähnt die Standards nicht.<sup>69</sup>

Eine Ausnahme sind Schiffler und Hübner, welche in einem kurzen Kapitel im *Handbuch Evaluation psychologischer Interventionsmaßnahmen* die JC-Standards als die „derzeit wohl [...] wichtigsten Evaluationsstandards“ bezeichnen.<sup>70</sup>

Im Kontext der internationalen Zusammenarbeit existieren anerkannte Regelwerke.<sup>71</sup> Hier sind zunächst die in ihrer Ursprungsfassung von 1991 sechsseitigen „*DAC-Principles of Development Assistance*“ zu nennen, die sich auf Evaluation in der Entwicklungszusammenarbeit beziehen und inzwischen in den Mitgliedsländern der OECD weitgehend akzeptiert sind.<sup>72</sup> Sie fungieren zwar

<sup>55</sup> Vgl. Jang (2000).

<sup>56</sup> Z. B. im Standard K4 „Human gestaltete Interaktionen“ (*Human Interactions*), dessen erste Richtlinie fordert: „Die Evaluatorinnen sollten sich intensiv bemühen, kulturelle Hintergründe, soziale Werte und Sprachunterschiede der Teilnehmenden zu verstehen.“ Joint Committee (2000), 127.

<sup>57</sup> Vgl. Nowacki (2000).

<sup>58</sup> Vgl. Sanders (2000).

<sup>59</sup> Persönliche Konversation der Autorin mit Daniel Stufflebeam und James Sanders, zwei ehemaligen Vorsitzenden des Joint Committee, im April 2000.

<sup>60</sup> Evaluation Practice News (Juni 1994).

<sup>61</sup> Im international mit Titeln zur Evaluation wohl führenden Verlag Sage (Thousand Oaks CA, London, New Delhi) rangieren sie auf den ersten beiden Plätzen der Evaluations-Bestseller (telefonische Auskunft eines Sage-Mitarbeiters, London (September 2000).

<sup>62</sup> Rossi, Freeman, Lipsey (1999), 426.

<sup>63</sup> Patton (1997), 16f.

<sup>64</sup> Owen, Rogers (1999), 153ff.

<sup>65</sup> Siehe URL <http://www.parklane.com.au/aes/ethics.htm> [September 2000]

<sup>66</sup> Villa (1998) bzw. Beywl, Widmer (2000).

<sup>67</sup> Bortz, Döring (1995).

<sup>68</sup> Z. B. Wottawa, Thierau (1998).

<sup>69</sup> Vedung (1999), Originale 1991/1997. Dies sind Indizien dafür, dass der Paradigmenwandel, der in der amerikanischen Evaluation seit Mitte der 70er Jahre stattgefunden hat (vgl. Beywl 1988; aktuell Stufflebeam 1999), bislang auf dem alten Kontinent vielfach in Lehrbüchern in seiner Bedeutung nicht verarbeitet ist.

<sup>70</sup> Schiffler, Hübner (2000), in: Hager, Patry, Brezing (2000), 141.

<sup>71</sup> Eine ausführlichere, in einigen Punkten weniger aktuelle Darstellung findet sich in Beywl, Widmer (2000). Dort wird eine Analyse der verschiedenen Standard-Werke bezüglich ihres inhaltlichen Schwerpunktes, der Qualitätsträger, der Regelsetzer, des Sanktionsgehaltes und des Umfanges geleistet.

<sup>72</sup> OECD (1998).



auch als Programmevaluationsstandards, formulieren also Qualitätsanforderungen an Evaluationen, sind aber überwiegend Programmstandards, enthalten also Anforderungen an die zu evaluierenden Programme selbst. Laut Nowacki lassen sich trotz der unterschiedlichen allgemeinen Zielsetzung der beiden Regelwerke in vielen der DAC-Richtlinien die Standards des Joint Committee (implizit) wiederfinden.<sup>73</sup>

Die PUMA-Guidelines sollen Mitgliedsstaaten der OECD dabei unterstützen, die Nutzung von Evaluation im Rahmen von Performance-Management-Systemen zu verbessern.<sup>74</sup> Sie wenden sich vorrangig an die für die politische Steuerung von Evaluation Verantwortlichen (Regierungsorganisationen, Politiker und leitende Beamte). Insgesamt werden neun Richtlinien mit jeweils zwei bis fünf thesenartig gefassten Absätzen formuliert. Bei zahlreichen inhaltlichen Übereinstimmungen mit den JC-Standards privilegieren die PUMA-Guidelines die Nutzungsart ‚Entscheidungsvorbereitung‘ deutlich gegenüber anderen und schaffen eine erhebliche Spannung zwischen Entscheidungsorientierung und ihrem explizit formulierten partizipativem Anspruch.

Das MEANS-Handbuch No. 5 (MEANS steht für ‚*Methods for Evaluating Actions of a Structural Nature*‘ und resultiert aus einem Förderprogramm der Europäischen Kommission, 1996) stellt eine 42-seitige Broschüre im Format A5 dar, die zur Qualität von Evaluationen im Rahmen von Europäischen Strukturprogrammen beitragen soll. Das Handbuch enthält einen expliziten Hinweis auf die JC-Standards, übernimmt viele ihrer Anforderungen, hat aber insgesamt einen schmaleren Fokus. Es bestätigt die Relevanz der mit den JC-Standards angesprochenen Dimensionen für die Evaluation der Strukturfondsmaßnahmen der Europäischen Union.

Im Januar 2000 veröffentlichte die Europäische Kommission die *Good Practice Guidelines for the Management of the Commission's Evaluation Function* (GPGEF).<sup>75</sup> Als Standards der Evaluation werden genannt und mit je wenigen Sätzen beschrieben: Relevanz für die Adressaten, Zugang zu relevanten Informationen, Identifikation und Einbeziehung von Beteiligten und Betroffenen, Zuverlässigkeit der Datenerhebungen sowie Transparenz und Objektivität. In der Fortschreibung wird als wichtiger Grundsatz aufgenommen: „*As a general rule, evaluation results should be made publicly available.*“ Dieser letztgenannte Grundsatz ist ein großer Schritt in Richtung freien Informationszugangs zu Evaluationsergebnissen, der den Bürgern und Bürgerinnen der EU zugestanden wird, und geht deutlich über die vorsichtigeren Anforderungen der JC-Standards hinaus. Diese Guidelines des *Network of Evaluators* bei der Europäischen Kommission sind sehr knapp formuliert, weisen einige Überschneidungen mit den JC-Standards auf und sind in Gliederung und Darstellung vergleichsweise unsystematisch.

#### 4. Ausblick

Die Entwicklung und Nutzung von Standards der Evaluation steht auf den Tagesordnungen von Regierungsorganisationen, von großen sozialen Organisationen,<sup>76</sup> privaten Unternehmen sowie der Evaluatoren und Evaluatoreninnen. In welchem Maße ein zentraler Bezugspunkt, wie z. B. die JC-Standards, gewählt wird oder nationale, politikfeldspezifische bzw. disziplinäre Standards nebeneinander bestehen oder mit einem übergreifenden Regelwerk verbunden werden, ist eine spannende Frage der Diskussion in den kommenden Jahren.

Die Schweizerische Evaluationsgesellschaft (SEVAL) hat mit den von ihr empfohlenen Standards, die eng an die des Joint Committee angelehnt sind, vergleichsweise schnell ein Regelwerk entwickelt. Dahinter steht eine mehrjährige Förderung der Entwicklung von Evaluations-Know-How durch das Nationale Forschungsprogramm „Wirksamkeit staatlicher Maßnahmen“ (NFP 27), in dessen Rahmen beispielhafte Evaluationen realisiert und umfassende Metaevaluationen durchgeführt wurden<sup>77</sup> sowie Lehrbücher entstanden<sup>78</sup>. Die von der SEVAL erarbeiteten Standards sind darauf angelegt, sowohl für Auftraggeber als auch Auftragnehmer in der Evaluation eine fachliche Basis zu bieten, und zwar übergreifend für alle Bereiche der öffentlichen Politik und Verwaltung sowie privater und gemeinnütziger Organisationen, sowohl bezogen auf große, im nationalen Rahmen durchgeführte Programme als auch auf kleine, wie etwa das Schulprogramm eines Gymnasiums. Die JC-Standards in der Buchfassung (englisch und deutsch) bieten für die SEVAL-Standards einen breiten fachlichen Hintergrund, der für Beauftragung, Planung, Durchführung und Evaluation von Evaluationen konsultativ genutzt werden kann.

Die bisher entwickelten Regelwerke im internationalen und europäischen Raum (DAC, PUMA, MEANS, GPGEF) stehen kaum in Bezug zueinander. Letztgenannte sind meist im Alleingang oder doch überwiegend durch die internationalen Behörden erstellt und nehmen die reife Methodologie der Evaluation, wie sie sich in aktuellen Lehrbüchern oder auch den JC-Standards findet, nicht immer

<sup>73</sup> Nowacki (2000a).

<sup>74</sup> OECD — Public Management Service (1998).

<sup>75</sup> Siehe <http://europa.eu.int/comm/budget/evaluation/en/goodpractice.htm> [September 2000]

<sup>76</sup> Beispielsweise stellt die Internationale Föderation der Rotkreuz- und Rothalbmondgesellschaften (2000, 15) in ihrer Strategie 2010 heraus, wie wichtig es für die gesamte Föderation ist, Evaluationssysteme einzuführen: „So sollen Fortschritte bei den Kernaufgaben messbar werden und Lernprozesse in künftige Entwicklungsprogramme von Beginn an aufgenommen werden können.“

<sup>77</sup> Vgl. Widmer (1996).

<sup>78</sup> Z. B. Bussmann et al. (1997).

in befriedigendem Maße auf. So ist die teils vorfindbare intransparente Vermischung von Programmstandards (z. B. Gütekriterien für Programme) und Evaluationsstandards (Qualitätsmerkmale für Evaluation) problematisch. Dabei zeigt sich auch in diesen Regelwerken das Bestreben, politikfeldübergreifend gültige, interkulturell anwendbare und interdisziplinär orientierte Regelwerke zu schaffen. Die GPGEF-Verpflichtung auf Veröffentlichung von Evaluationsberichten weist darauf hin, dass Evaluation für politische Kultur und gelebte Demokratie große Potentiale aufweist.

Die „Deutsche Gesellschaft für Evaluation“ (DeGEval) beschäftigt sich seit 1998 mit der Frage, ob Standards der Evaluation und / oder an Evaluatorinnen und Evaluatoren gerichtete Leitprinzipien erarbeitet werden sollen. Im Herbst 2000 fand eine Mitgliederbefragung statt, in der es um Anforderungen der Mitglieder und ihre Präferenzen für ein solches Regelwerk geht. Die meisten antwortenden Mitglieder bevorzugen — vor die Alternative ‚knappe Leitprinzipien‘ oder ‚ausführlichere Standards‘ gestellt<sup>79</sup> — ein umfassenderes Papier.<sup>80</sup>

Über diese Grundsatzentscheidung hinaus werden u. a. folgende Fragen zu klären sein:

- Wer ist Träger des Regelwerkes? Die Deutsche Gesellschaft für Evaluation als Zusammenschluss von Auftraggebern und Auftragnehmern von Evaluationen? Oder ein Komitee, in dem ähnlich dem Joint Committee zahlreiche Fachorganisationen und Berufsverbände zusammen arbeiten?<sup>81</sup>
- Werden Auftraggeber, Auftragnehmer und weitere Gruppierungen (z. B. fachwissenschaftliche Gesellschaften, Berufsverbände) an der Entwicklung der Standards beteiligt; in welcher Form soll dies geschehen? Wie beteiligt sich die öffentliche Hand als Hauptauftraggeber von Evaluationen an diesem Prozess? Durch Personen, die sie entsendet oder in institutioneller Form?<sup>82</sup>
- Wird Interdisziplinarität betont; gelingt es, die in Deutschland weiterhin stark disziplinär organisierten Fachwissenschaften einzubinden? Welche Vorstellungen bringen, neben Sozialwissenschaftlern, z. B. auch Rechtswissenschaftler oder Betriebswirte / Ingenieure in den Standardsetzungsprozess ein?
- Strebt das Papier Gültigkeit über breite Anwendungsbereiche und Politikfelder an oder sollen einige wegen bestehender Besonderheiten ausgenommen werden; oder soll es für einige Bereiche evtl. später spezifische Zusatzregelungen geben?<sup>83</sup>
- Wird der Anschluss hergestellt an internationale / im internationalen Raum verbreitete Standards; geschieht dies durch Hinweise und Quellennachweise oder darüber hinaus gehend durch Übernahme der Struktur bestehender Regelwerke? Wird von vornherein eine Har-

monisierung von nationalen Standards im europäischen Raum oder darüber hinaus angestrebt?<sup>84</sup>

- Welchen Grad der Verbindlichkeit soll ein solches Regelwerk erlangen? Dies hängt stark davon ab, wer es trägt und für wen welche Art von Verbindlichkeit, bis hin zu Sanktionen, sinnvoll und durchsetzbar ist.<sup>85</sup>

Zahl und Diversität der einbezogenen Organisationen und Verbände, Zeitrahmen der Erarbeitung, Finanzres-

<sup>79</sup> Als exemplarische Beispiele wurden die von der SEVAL empfohlenen Standards und die nur eine Druckseite umfassenden „Guidelines for Ethical Conduct“ der Canadian Evaluation Society zur Auswahl gestellt.

<sup>80</sup> Die ausführlichen Ergebnisse der Befragung sind auf der Homepage der DeGEval einsehbar (siehe URL: <http://www.degeval.de> [September 2000]).

<sup>81</sup> Sollen die Standards politikfeldübergreifend gelten, dürfte es schwierig sein, ein genügend kleines, arbeitsfähiges Komitee zu gründen.

<sup>82</sup> Die Bundes- und Landesregierungen verfügen in aller Regel nicht über zentrale Ansprechpartner für Evaluation; auch die meisten Ministerien haben keine Evaluations-Einheit (wie diese für alle EU-Generaldirektionen und auch nachgeordnete Dienststellen durchgängig eingeführt wird). Der Bundesrechnungshof beginnt, der Evaluation von Programmen einen eigenen Stellenwert bei der Prüfung des Regierungshandels zuzumessen; ein Indiz findet sich z. B. im Bericht für das Haushaltsjahr 1999 unter Pos. 29 „Förderung von Vermittlungsagenturen für Langzeitarbeitslose und andere schwervermittelbare Arbeitslose“ (Kapitel 11 12 Titel 685 01). Bei Fortführung der bestehenden Projekte hält der Rechnungshof „eine projektbegleitende, seine Erkenntnisse berücksichtigende Evaluation für notwendig.“ (Jahresbericht 1999, siehe <http://www.bundesrechnungshof.de/bem1999> [September 2000]).

<sup>83</sup> Als zwei Beispiele verwiesen sei auf die „Leitlinien zur Selbstevaluation — Ist unsere Schule auf dem richtigen Weg?“ (2000), veröffentlicht durch das Institut für Schulentwicklungsforschung an der Universität Dortmund und die Bertelsmann-Stiftung (<http://www.ifs.uni-dortmund.de>), die „Empfehlungen für die Evaluation in der Jugendhilfe“ (2000) sowie die gemeinsamen Empfehlungen von DPRG / GPRA (2000) zur Evaluation von Evaluationen im Bereich von PR-Maßnahmen.

<sup>84</sup> Ein Beispiel für die noch weit verbreitete Unentschiedenheit in der deutschen Diskussion ist das Gutachten der Präsidentin des Bundesrechnungshofes als Bundesbeauftragter für die Wirtschaftlichkeit in der öffentlichen Verwaltung. Von Wedel (1998) wählt als Titel des Gutachtens den Terminus „Erfolgskontrolle“ und verweist auf seine begriffliche Diffusität (S. 14). Im Abschnitt über „Erfolgskontrolle im internationalen Bereich“ hingegen wird der auch in der Internationalen Organisation der Rechnungskontrollbehörden genutzte Terminus „Evaluation“ mit „Erfolgskontrolle“ synonym gesetzt. Der Abschnitt schließt mit der Aussage „... dass Evaluationen oder Erfolgskontrollen als ein Steuerungselement angesehen werden, das die Übertragung größerer Verantwortlichkeiten auf dezentrale Verwaltungseinheiten erleichtern kann“ (S. 42).

<sup>85</sup> Die meisten Regelwerke zur Evaluation verzichten auf Sanktionen und tragen stärker den Charakter von Selbstverpflichtungen. Allerdings stellen fachöffentliche Erörterungen von — anonymen — Fällen in elektronischen Diskussionslisten, wie in der amerikanischen EVALTALK geführt, bereits eine erste und vielleicht oft wirksame Form der Kontrolle dar. Hinzu werden sicher Meta-Evaluationen treten, die durchgeführte Evaluationen im Nachhinein — evtl. auch vergleichend — bewerten.

sourcen und Stundenvolumen der im Hauptamt an der Entwicklung von Standards Mitarbeitenden, Grad der Professionalisierung der Beteiligten etc. stellen wichtige Rahmenbedingung für die Entwicklung von Standards der Evaluation dar. Es wird nicht einfach sein, bei gegenwärtig begrenzten Ressourcen für die Standardsetzung in der Evaluation einen befriedigenden Kompromiss zwischen zügiger Erarbeitung und Verabschiedung einerseits und hinreichender Passung auf die Erfordernisse der ver-

schiedensten Evaluationsfelder bei starker Konkretisierung der einzelnen Standards andererseits zu finden.

Schiebt man den Prozess der Erarbeitung und Verabschiedung von Standards zeitlich auf, läuft insbesondere die in Deutschland und Österreich junge Evaluationsprofession Gefahr, immer wieder mit schlechten Evaluationen und der öffentlichen Kritik daran konfrontiert zu sein, ohne präzisiert und schriftlich angeben zu können, welche Anforderungen gute Evaluationen erfüllen sollen.

## Literaturverzeichnis

- Berry, John W., Marshall H. Segall, Cigdem Kagıtcıbası (Hrsg.) (1997): Handbook of Cross-Cultural Psychology. Vol. 1 & 3, Needham Heights.
- Beywl, Wolfgang (1988, Reprint 1999): Zur Weiterentwicklung der Evaluationsmethodologie. Grundlegung, Konzeption und Anwendung eines Modells der responsiven Evaluation, Frankfurt.
- Beywl, Wolfgang (1996): Anerkannte Standards und Leitprinzipien der amerikanischen Evaluation. In: Heiner, Maja (Hrsg.), Qualitätsentwicklung durch Evaluation, Freiburg im Breisgau.
- Beywl, Wolfgang (1999): Standards für Evaluation. Auf dem Weg zu Regelwerken in der deutschsprachigen Evaluation. DeGEval News, Nr. 2(1), 1–4.
- Beywl, Wolfgang, Thomas Widmer (2000): Die ‚Standards‘ im Vergleich mit weiteren Regelwerken zur Qualität fachlicher Leistungserstellung. In: Joint Committee on Standards for Educational Evaluation, Handbuch der Evaluationsstandards, Opladen, 259–295.
- Blanke, Bernhard, Stefan v. Bandemer, Frank Nullmeier, Göttrik Wewer (Hrsg.): Handbuch zur Verwaltungsreform, Opladen.
- Bussmann, Werner, Ulrich Klöti, Peter Knoepfel (1997): Einführung in die Politikevaluation, Zürich.
- Caracelli, Valerie J., Leslie J.C. Riggın (1994): Mixed-method Evaluation: Developing Quality Criteria Through Concept Mapping. Mixed-method Collaboration. Evaluation Practice, Nr. 15(2), 139–52.
- DPRG (Deutsche Public Relations-Gesellschaft), GPRA (Gesellschaft Public Relations Agenturen (2000): PR-Evaluation. Messen — Analysieren — Bewerten. Bonn.
- „Empfehlungen für die Evaluation in der Jugendhilfe“ (2000). In: neue praxis. Zeitschrift für Sozialarbeit, Sozialpädagogik und Sozialpolitik, 30. Jhg. 2000, Heft 2, 190–192.
- European Union — Regional Policy and Cohesion (1996): MEANS-Handbook Nr. 5. Quality Assessment of Evaluation Reports: A Framework, Bruxelles.
- Grohmann, Romano (1997): Das Problem der Evaluation in der Sozialpädagogik. Bezugspunkte zur Weiterentwicklung der evaluationstheoretischen Reflexion, Frankfurt.
- Hager, Willi, Jean-Luc Patry, Hermann Brezing, (2000): Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien, Bern.
- Haller, Sabine. (1998): Beurteilung von Dienstleistungsqualität. Dynamische Betrachtung des Qualitätsurteils im Weiterbildungsbereich, 2. Auflage, Wiesbaden.
- Heinz, Rainer (2000): Kommunales Management. Überlegungen zu einem KGSt-Ansatz, Stuttgart.
- Hofstede, Geert (1980): Culture's Consequences. International Differences in Work-Related Values, Beverly Hills.
- Internationale Föderation der Rotkreuz- und Rothalbmondgesellschaften (2000): Strategie 2010. Das Leben von Menschen in Not und sozial Schwachen durch die Kraft der Menschlichkeit verbessern, Bonn (DRK Generalsekretariat).
- Jang, Soojung (2000): The Appropriateness of the Joint Committee Standards in Non-Western Settings: A Case Study of South Korea. In: The Program Evaluation Standards in International Settings. Occasional Paper Series, Nr. 17, The Evaluation Center, Western Michigan University, Kalamazoo, 41–59.
- Joint Committee on Standards for Educational Evaluation (1988): The Personnel Evaluation Standards, Newbury Park.
- Joint Committee on Standards for Educational Evaluation (2000): Handbuch der Evaluationsstandards, 2. Auflage, Opladen. (Original: The Program Evaluation Standards (1994), Thousand Oaks).

- Kaplan, Robert S; David Norton* (1997): *Balanced Scorecard*, Stuttgart.
- Lewy, Arieh* (1984): *Evaluation Standards: Comments from Israel*. *Evaluation and Program Planning*, Nr. 7, 377–381.
- Marklund, Sixten* (1984): *Applicability of Standards for Evaluations of Educational Programs, Projects, and Materials in an International Setting*. *Evaluation and Program Planning*, Nr. 7, 355–362.
- Nevo, David* (1984): *Applying the Evaluation Standards in a Different Social Context*. *Evaluation and Program Planning*, Nr. 7, 383–386.
- Nowacki, Julia* (2000): *Zur Tragfähigkeit der ‚Standards für Programmevaluationen‘ in einem trikulturellen Kontext*. *Evaluation eines technischen Fortbildungsprogramms für berufstätige Lehrer in Peru*. Diplomarbeit im Fach Erziehungswissenschaft an der Universität zu Köln.
- Nowacki, Julia* (2000a): *Die ‚Standards für Programmevaluationen‘ des ‚Joint Committee on Standards for Educational Evaluation‘ in der Entwicklungszusammenarbeit; Papier präsentiert auf der 4. Sitzung des Arbeitskreises Evaluation von Entwicklungspolitik in der Deutschen Gesellschaft für Evaluation; am 11./12. Mai 2000 in Kehl*. URL: [http://www.degeval.de/ak\\_ez/ak\\_ez4.pdf](http://www.degeval.de/ak_ez/ak_ez4.pdf) [Stand September 2000].
- Nullmeier, Frank* (1998): *Input, Output, Effektivität und Effizienz*. In: *Blanke, Bernhard, Stefan v. Bandemer, Frank Nullmeier, Göttrik Wewer* (Hrsg.): *Handbuch zur Verwaltungsreform*. Opladen, 314–322.
- OECD* (1998): *Review of the DAC Principles for Evaluation of Development Assistance*, Paris.
- OECD — Public Management Service* (1998): *Best Practice Guidelines for Evaluation*, Paris.
- Owen, John M., Patricia J. Rogers* (1999): *Program Evaluation: Forms and Approaches*, Thousand Oaks.
- Patton, Michael Q.* (1997): *Utilization-Focused Evaluation*. The New Century Text, Thousand Oaks.
- Rossi, Peter H., Howard E. Freeman, Mark W. Lipsey* (1999): *Evaluation. A Systematic Approach*. 6. Auflage, Thousand Oaks.
- Russon, Craig* (Hrsg.) (2000): *The Program Evaluation Standards in International Settings*. Occasional Paper Series, Nr. 17, The Evaluation Center, Western Michigan University, Kalamazoo.
- Russon, Craig, Mahesh Patel* (1999): *Appropriateness of the Program Evaluation Standards for Use in African Cultures*. Zur Publikation versendetes Manuskript.
- Sanders, James R.* (1997): *Cluster Evaluation*. In: *Chelimski, Eleanor, William R. Shadish* (Hrsg.), *Evaluation for the 21st Century. A Handbook*, Thousand Oaks, 396–404.
- Sanders, James R.* (2000): *Aspekte der Entwicklung und Verbreitung der Evaluationsstandards*. In: *Joint Committee on Standards for Educational Evaluation*, *Handbuch der Evaluationsstandards*, 2. Auflage, Opladen, 237–241.
- Schiffler, Angelina, Sigrid Hübner* (2000): *Allgemeine Standards für die Evaluationspraxis*. *Die Standards des ‚Joint Committee on Standards for Educational Evaluation‘ und ihre Anwendung auf praktische Aspekte bei der Evaluation von psychologischen Interventionsmaßnahmen*. In: *Hager, Willi, Jean-Luc Patry, Hermann Brezing* (Hrsg.), *Handbuch der Evaluation psychologischer Interventionen*, Bern, 141–151.
- Schwartz, Shalom H.* (1994a): *Beyond Individualism/Collectivism: New Cultural Dimensions of Values*. In: *Kim, Uichol, Harry C. Triandis, Cigdem Kagitcibasi, Gawon Yoon* (Hrsg.), *Individualism and Collectivism: Theory, Method and Applications*, Thousand Oaks, 85–119.
- Scriven, Michael* (1991): *Evaluation Thesaurus*. 4. Auflage, Newbury Park.
- Smith, Nick L., Savior Chircop, Prachee Mukherjee* (1993): *Considerations on the Development of Culturally Relevant Evaluation Standards*. *Studies in Educational Evaluation*, Nr. 19, 3–13.
- Stufflebeam, Daniel L.* (1986): *Standards of Practice for Evaluators*. Vortrag bei der Jahrestagung der American Educational Research Association, San Francisco.
- Stufflebeam, Daniel L.* (1999): *Foundational Models for 21st Century Program Evaluation*. Occasional Paper Series, Nr. 16, The Evaluation Center, Western Michigan University, Kalamazoo.
- Stockdill, Stacey H.* (1986): *Evaluation Standards. A Study of their Appropriateness in Business & Education*. Ph.D. University of Minnesota. *Dissertation Abstracts International (DAI)*, 47-08A:2967.
- Taut, Sandy* (2000): *The Cross-Cultural Transferability of the Program Evaluation Standards*. In: *The Program Evaluation Standards in International Settings*. Occasional Paper Series, No. 17, The Evaluation Center, Western Michigan University, Kalamazoo, 5–27.
- Triandis, Harry C.* (1995): *Individualism & Collectivism*, Boulder.
- Vedung, Evert* (1999): *Evaluation im öffentlichen Sektor*, Wien.
- Villa, Aurelio* (1998): *Estandares para la evaluacion de programas [Standards für Programmevaluation]*, Bilbao.
- v. Wedel, Hedda* (1998): *Erfolgskontrolle finanzwirksamer Maßnahmen in der öffentlichen Verwaltung (Schriftenreihe der Bundesbeauftragten für die Wirtschaftlichkeit in der Verwaltung, Bd. 2)*, Stuttgart.

- Wholey, Joseph S., Harry P. Hatry, Kathryn E. Newcomer* (1994): Handbook of Practical Program Evaluation, San Francisco.
- Widmer, Thomas* (1995): Evaluating Evaluations. Does the Swiss Practice Live up to The Program Evaluation Standards? (Paper presented at the 1. International Evaluation Conference, Vancouver, Canada, 1-5 November 1995). In: The Program Evaluation Standards in International Settings. Occasional Paper Series, No. 17, The Evaluation Center, Western Michigan University, Kalamazoo, 67–80.
- Widmer, Thomas* (1996): Meta-Evaluation. Kriterien zur Bewertung von Evaluationen, Bern.
- Widmer, Thomas* (2000): Kontext, Inhalt und Funktion der ‚Standards der Programmevaluation‘. In: Müller-Kohlenberg, Hildegard (Hrsg.), Evaluation und Qualitätsmanagement in der sozialen Arbeit, Opladen (im Erscheinen).
- Widmer, Thomas, Wolfgang Beywl* (2000): Die Übertragbarkeit der Evaluationsstandards auf unterschiedliche Anwendungsfelder. In: Joint Committee on Standards for Educational Evaluation, Handbuch der Evaluationsstandards, Opladen, 243–257.
- Widmer, Thomas, Charles Landert, Nicole Bachmann* (1999): Evaluations-Standards empfohlen von der Schweizerischen Evaluationsgesellschaft (SEVAL), Geneve.
- Wollmann, Hellmut* (2000): Staat und Verwaltung in den 90er Jahren. Kontinuität oder Veränderungswelle? In: Czada, Roland, Helmut Wollmann (Hrsg.), Von der Bonner zur Berliner Republik, Leviathan Sonderheft, Nr. 19, 694–731.
- Worthen, Blaine R., James R. Sanders, Jody L. Fitzpatrick* (1997): Program Evaluation. 2. Auflage, White Plains.
- Wottawa, Heinrich, Heike Thierau* (1998): Lehrbuch Evaluation, 2. Auflage, Bern.
- Wulf, Christoph* (Hrsg.) (1992): Evaluation. Beschreibung und Bewertung von Unterricht, Curricula und Schulversuchen, München.

## Standards: A Strategy For Improving Quality in Evaluation

### Summary

*There is growing demand for evaluations in the EU as well as in Germany. Evaluation has developed independently from approaches such as "New Public Management" or "Quality Management". Since Evaluation is rooted in human services, it has special significance in those fields of public policy which have to rely on active cooperation of citizens. The American "Program Evaluation Standards" (JC-Standards) that originate in the field of education serve as an example to delineate how high quality of evaluations can be secured. The article clarifies the meaning and use of "program" as a technical term in the field of evaluation and discusses the structure and contents of the JC-Standards. The authors provide evidence for their applicability in a broad range of public policy fields and present demands on the Standards' cross-cultural transferability. Recent references serve to test the overall significance of the JC-Standards, followed by an analysis of the possibilities for their integration in new national and international guidelines for evaluation. The article closes with a call for the acceleration of the standard setting process in Germany and Europe.*