

Die Key Evaluation Checklist

Michael Scriven

von Daniela Schröter und Wolfgang Beywl 2006 übersetzt und kommentiert,

basierend auf alter Original-Version

Original von 2007: http://www.wmich.edu/evalctr/archive_checklists/kec_feb07.pdf

Original (2007 überarbeitet):

http://www.wmich.edu/sites/default/files/attachments/u350/2014/key_evaluation_checklist.pdf

Vorbemerkungen der Übersetzerin und des Übersetzers	2
TEIL A: PRÄLIMINARIEN	6
I. Management Summary	6
II. Vorwort	6
III. Methodologie	8
TEIL B: GRUNDLAGEN	8
1. Hintergrund und Kontext	8
2. Beschreibungen und Definitionen	9
3. Konsumierende « <i>impactees</i> »	9
4. Ressourcen (auch Stärkenabschätzung « <i>strengths assessment</i> »)	10
5. Werte	10
TEIL C: UNTER-EVALUATIONEN	13
6. Prozess	13
7. Outcomes	13
8. Kosten	14
9. Vergleiche	15
10. Generalisierbarkeit	15
TEIL D: SCHLUSSFOLGERUNGEN	16
11. Gesamtbedeutung	16
12. Fakultative Empfehlungen und Erklärungen	16
13. Fakultative Verantwortung und Rechenschaftslegung	17
14. Bericht und Unterstützung	18
15. Meta-Evaluation	19
<i>Literatur</i>	20
<i>Disclaimer</i>	21



Vorbemerkungen der Übersetzerin und des Übersetzers

Die *Key Evaluation Checklist (KEC)*¹ wurde von Michael Scriven in den frühen 1970er Jahren für „*Educational Testing Services*“ (ETS) entwickelt, um Bildungsprodukte wie Curricula oder Unterrichtsmaterialien, die von bundesfinanzierten Forschungs- und Entwicklungszentren erstellt worden waren, zu evaluieren. Nach ihrer Erstveröffentlichung (Scriven mit Roth 1977) im *Evaluation Thesaurus*² wurde die Checkliste mehrfach überarbeitet.³ Es handelt sich nicht um eine übliche Checkliste mit kurzen Merkpunkten oder Anweisungen, die – falls zutreffend – abgehakt «*checked*» werden, sondern um einen in 15 Abschnitte („Checkpunkte“) gegliederten Text, der Evaluierende bei der professionellen Umsetzung ihrer Aufgaben anleitet.

Die „Checkpunkte“ bereiten in über drei Jahrzehnten gewonnene intensive Erfahrungen zum Evaluationshandeln auf. Die KEC ist ausser für Evaluationsnovizen speziell nutzbringend für Personen, die schon lange in der Evaluation tätig sind und einen Gegencheck machen wollen, ob sie wirklich nichts Wichtiges vergessen und ob sie ihr Vorgehen gründlich durchdacht haben. Der Text ist sehr dicht geschrieben, da er vieles auf gedrängtem Raum ansprechen will, und teilweise stichwortartig formuliert. Gleichzeitig gibt es zahlreiche Querbezüge zu Evaluationsmodellen und Grundkonzepten der Evaluation, die sich v. a. Personen mit intimer Kenntnis einer breiten Evaluationsliteratur erschliessen. Darauf weist Scriven in seiner ersten „Hauptanmerkung“ ausdrücklich hin.

Die KEC stellt aus Scrivens Sicht einen Rahmen dar, in welchem diverse Evaluationsmodelle verwendet werden können.

Kennzeichnend für die Logik der KEC ist es, vom Abschluss der Evaluation her zu denken. Daher beginnt die Checkliste mit Ausführungen zum Bericht. Die Checkliste ist Ausdruck der Anforderung, dass ein Evaluationsplan durch Iteration an Substanz gewinnt – ein in der Regel mehrmaliges, gedankliches Vor- und Zurückgehen zwischen den Checkpunkten ist gefordert. Die Checkpunkte sind in der Realität der Evaluationsplanung und Evaluationsteuerung stark voneinander abhängig und betreffen überschneidende Themen:

Die drei voran gestellten „Hauptanmerkungen“ enthalten eine Definition von Evaluation und weiterer Grundbegriffe, umreissen die Arten von Evaluationsgegenständen, für welche die KEC gilt, und betonen den Grundsatz des iterativen Vorgehens.

Teil A – Präliminarien - (I. Management Summary, II. Vorwort, III. Methodologie), behandelt drei grundlegende Elemente von Evaluationsberichten, die über die gesamte Evaluation hinweg beachtet werden sollen: Während die Management Summary eine allgemeine kurze, ein- bis zweiseitige Zusammenfassung der abgeschlossenen Evaluation bietet, werden im Vorwort Details zum Hintergrund der Evaluation beschrieben (z. B. Wer hat die Evaluation angefordert und warum?). Im Abschnitt zur Methodologie wird insbesondere auf die wissenschaftliche und

-
- 1 Soweit wir englischsprachige Termini zusammen mit der deutschen Übersetzung nutzen, setzen wir sie in doppelte spitze Klammern und zeichnen sie kursiv aus (z. B. «*merit*»).
 - 2 *Evaluation Thesaurus*, Erstausgabe mit Jane Roth, Edgepress, April 1977. dritte Ausgabe (Einzelautor *Michael Scriven*), 1982, 6 Nachdrucke. Vierte Ausgabe, Sage, 1991; Fünfte Ausgabe in Vorbereitung.
 - 3 Dieser Übersetzung liegt die Fassung der KEC von 23. Oktober 2005 zugrunde: http://www.wmich.edu/evalctr/checklists/kec_october05.pdf [Stand: 02.06.2006]. Neue Version hier. Im Laufe der im Zusammenhang mit der Übersetzung geführten Diskussionen hat Michael Scriven einige Ergänzungen zu dieser Version gemacht, die bereits mit übersetzt sind. Es handelt sich insbesondere um Ausführungen im Abschnitt „Werte“ – vgl. Fussnote 20.



Wissensbasis eingegangen und es wird geklärt, welche Evaluationsmodelle, Untersuchungsdesigns und Methoden verwendet werden.

Teil B – Grundlagen - behandelt die über den Evaluationsgegenstand zu gewinnenden Informationen (Fakten): Warum wurde der Evaluationsgegenstand ins Leben gerufen (1. Hintergrund und Kontext)? Was ist und tut der Evaluationsgegenstand (2. Beschreibungen und Definitionen)? Wer sind die direkt Konsumierenden (Teilnehmende) und potenziell global beeinflussten Personengruppen des Evaluationsgegenstandes (3. Konsumierende/Beeinflusste «*impactees*»)? Welche Ressourcen tragen zur Realisierung des Evaluationsgegenstandes bei oder könnten zur Realisierung beitragen (4. Ressourcen/Stärkenabschätzung)? Auf welcher Basis kann beurteilt werden, ob der Evaluationsgegenstand von hohem kostenfreiem bzw. kostenabhängigem Wert oder von hoher Bedeutung ist (s. 5. „Werte“)?

In Teil C – Unter-Evaluationen - werden die in Checkpunkt 5 festgelegten Werte und Standards mit den „Fakten“ in die entsprechenden Unter-Evaluationen «*sub-evaluations*» integriert, um evaluative Ergebnisse zu erzielen. Die Werte können auch der Entwicklung von Kriterien dienen, woraus Indikatoren zur Datenerhebung abgeleitet werden können. Wie gut, wertvoll, effizient ist der Evaluationsgegenstand in seiner Art und Weise (6. Prozesse)? Wie gut und wertvoll sind intendierte und nicht-intendierte Auswirkungen auf Konsumierende und andere beeinflusste Personengruppen (7. Outcomes)? Wie effektiv ist der Evaluationsgegenstand hinsichtlich monetärer und nicht-monetärer Kosten (Checkpunkt 8) und im Vergleich zu Alternativen (Checkpunkt 9)? Wäre es sinnvoll, wertvoll und/oder bedeutend, den Evaluationsgegenstand oder Elemente davon in andere Kontexte zu exportieren, zu übertragen oder zu transportieren? Hat der Evaluationsgegenstand Nachhaltigkeit? Ist er haltbar, dauerhaft und/oder belastbar (10. Generalisierbarkeit)?

Teil D – Schlussfolgerungen - beschreibt den Weg zu diesen Endergebnissen der Evaluation. Unter-Evaluationen werden zusammengefasst und die übergreifenden Fragen der Evaluation beantwortet (11. Gesamtbedeutung), mögliche Empfehlungen und Erklärungen (Checkpunkt 12) geliefert und je nachdem werden Verantwortung und Rechenschaftslegung zugeschrieben (Checkpunkt 13). Danach wird geklärt, wie die Berichterstattung erfolgt, wem berichtet werden soll und in welcher Form die Berichterstattung stattfindet sowie welche Formen der Unterstützung angeboten werden (14. Bericht und Unterstützung). Im letzten Checkpunkt (15.) wird auf die Meta-Evaluation eingegangen, welche als Selbstkontrolle dienen kann, doch idealerweise von extern Evaluierenden durchgeführt wird. Die Meta-Evaluation kann sowohl anhand der KEC als auch mit Hilfe anderer Standards durchgeführt werden. Generell sollten alle aufgeführten und nicht als *fakultativ* gekennzeichneten Checkpunkte in der Evaluation eingeschlossen sein, um Validität zu erzielen (Siehe *Davidson*, 2005).

Die Übersetzungstätigkeit der KEC unterscheidet sich dramatisch von der bei anderen Checklisten oder ‚normalen Lehrbuchtexten‘. Sie ähnelt mehr der Übertragung eines literarischen Textes von der Original- in eine Fremdsprache. Die KEC ist wegen der hohen Dichte des Textes, den vielen, teils impliziten Querbezügen und den manchmal fast unmerklich, manchmal völlig unerwartet eingestreuten humorvollen und ironischen Anspielungen kein offenes Buch. Viele zunächst verblüffende Bemerkungen sind auf den transdisziplinären Charakter der Evaluation zurückzuführen und die Tatsache, dass Scriven alltägliche Bewertungen (er nennt sie ebenfalls „Evaluationen“) nicht abwertet, sondern in ihnen lehrreiche Beispiele für professionelle Evaluationen sieht. So verweist er auf eine Vielzahl von Evaluationsbranchen (u. a. Produkt-, Personal-, Leistungs-, Angebote-, Portefeuille-, Politik- und Programmevaluation), Evaluationen innerhalb aller konventionellen Disziplinen und Fachrichtungen



(u. a. in der Kriminalistik durch forensische Fachpersonen, bei Gerichtsverhandlungen, in den Sozial- und Verhaltenswissenschaften, in den Geisteswissenschaften) sowie auf Evaluationen in verschiedensten Lebenslagen (z. B. bei Wettbewerben, Sportveranstaltungen, beim wöchentlichen Einkauf, bei der Wohnungssuche usw.).

Teils haben wir den Eindruck, der Text – wie immer übersetzt – erfordert fortlaufend Decodierungen und Interpretationen aus der Perspektive der Lesenden. Sein Bedeutungsgehalt wird sich mit jeder weiteren ‚Iteration‘ verändern. Die Übersetzung ist daher eine Gratwanderung zwischen den beiden Möglichkeiten, mehrdeutige Passagen entweder (einseitig) aufzuklären und ihren Bedeutungsgehalt (und ihr reflektionsförderndes Potenzial) damit einzuengen oder sie in Ihrer Offenheit darzubieten und sie damit für manche Lesende verwirrend und verunsichernd zu belassen. Wir hoffen, die Balance auf dem schmalen Grat gehalten zu haben, den Diskursbereiten Varianten der Interpretation offen lassend, den Sicherheit Suchenden genügend Klarheit und Nachvollziehbarkeit bietend.⁴

Bei der Übersetzung wurde darauf geachtet, grammatikalische Bezüge (z. B. von Relativpronomen) explizit zu machen, etwa indem das Bezugswort (im Unterschied zu Originalfassung) wiederholt wird. Dadurch und durch die Übertragung ins Deutsche ist die übersetzte Fassung deutlich länger als die Originalfassung.

Im Verlauf der Übersetzung haben wir mehrfach deutungsbedürftige Punkte mit Michael Scriven diskutiert, wofür wir ihm herzlich danken.⁵ Auf diesem Weg ist es gelungen, den Text nahe an den angestrebten Bedeutungen des Urhebers in eine andere Sprache zu übersetzen. Dabei ist es unvermeidlich, dass Konnotationen in den anderen evaluationskulturellen Kontext der deutschsprachigen Länder nicht gesichert übertragbar sind und somit die Übersetzung und deren Rezeption durch die Lesenden (vorzugsweise in Deutschland, Österreich und der deutschsprachigen Schweiz) eigenständige Interpretationsleistungen darstellen.

An dieser Stelle muss auch festgehalten werden, dass sich Michael Scrivens Einstellungen gegenüber Evaluation und ihrer Umsetzung ständig weiterentwickeln. Somit wurde die KEC innerhalb der letzten zwei Jahre mehr als fünfmal überarbeitet. Der Begriff *ascriptive evaluation* wurde mit der Initiierung des *Journal of MultiDisciplinary Evaluation*⁶ eingeführt, *proformative evaluation* in einer E-Mail zum Evaltalk Listserver am 8. März, 2006 hinzugefügt. Des Weiteren werden in letzter Zeit auch *highbars* und *lowbars* diskutiert. Ein Nachteil dieser Entwicklungen liegt darin, dass die 4. Auflage von Scriven *Evaluation Thesaurus* aus dem Jahr 1991 nicht mehr aktuell ist, und dass eine neue und überarbeitete fünfte Auflage noch nicht zur Verfügung steht.

-
- 4 Als ergänzende Lektüre finden Sie eine klare und leicht verständliche Anleitung zur Realisierung einer Evaluation auf Basis der KEC bei Davidson (2005).
- 5 So z. B. die Übersetzung des Begriffspaares *merit/worth*, die Bedeutung von *impactees* oder den Bedeutungsumfang von *policy* in diesem Text
- 6 Scriven, Michael (2004). *Editorial: The Fiefdom Problem. Journal of MultiDisciplinary Evaluation* 1. Frei verfügbar unter: <http://evaluation.wmich.edu/jmde/> [Stand 02.06.2006]



Key Evaluation Checklist (KEC)

Michael Scriven

2005

Original: http://www.wmich.edu/evalctr/checklists/kec_october05.pdf

Stand des Originals: 31. 3. 2005 (im Checkpunkt 5 – Werte – von M. S. gegenüber der zugrunde liegenden Version vom 23. 10. 2005 weiterentwickelt)
von Daniela C. Schröter und Wolfgang Beywl 2006 übersetzt und annotiert.

Bestimmt für die Planung, Durchführung und Berichterstattung von Evaluationen, die sich auf Programme, Pläne und Politiken richten; ausserdem für die Evaluation von Evaluationen dieser Evaluationsgegenstände (Metaevaluation).

HAUPTANMERKUNG A: In diesem Dokument ist „Evaluation“ als die Bestimmung von kostenfreiem Wert «*merit*», kostenabhängigem Wert «*worth*», und Bedeutung «*significance*» eines Evaluationsgegenstandes definiert⁷. „Evaluationsgegenstand“ «*evalund*» meint all das, was evaluiert wird. Die Checkliste richtet sich an aktiv Evaluierende.⁸ Somit werden Kenntnisse evaluationsspezifischer Fachbegriffe wie z. B. „formative“, „zielfreie Evaluation“ «*goal-free evaluation*» oder „*Ranking*“ vorausgesetzt und nicht definiert. Sie können in Scrivens *Evaluation Thesaurus* (4. Auflage 1991) nachgeschlagen werden.⁹ Des Weiteren wird in der KEC der Einfachheit halber generisch von „Programmen“ gesprochen und nicht zwischen „Programmen, Plänen, Politiken, und Metaevaluationen sowie darauf gerichteten Evaluationsplänen oder Evaluationsberichten“ unterschieden.

HAUPTANMERKUNG B: Die KEC kann ausserdem bei sorgfältiger Anwendung für die Evaluation von Produkten und Organisationseinheiten (z. B. Abteilungen) verwendet werden, sowie bei hohen Anforderungen an die erforderlichen Transferleistungen, für einige Aufgaben der Personalevaluation.

-
- 7 *Merit* und *worth* werden im Deutschen – z. B. im Handbuch der Evaluationsstandards (Sanders 2006) oder in den Evaluationsstandards der *DeGEval* (2002) – oft mit „Güte“ und „Verwendbarkeit“ übersetzt. Dieses Wortpaar setzt üblicherweise den intrinsischen Wert eines Evaluationsgegenstandes seinem kontextabhängigen Wert gegenüber (vgl. Beywl 1988). Scriven verschiebt mit seiner Abgrenzung der beiden Begriffe die Trennlinie zwischen ihnen (sein *merit* enthält auch bereits Aspekte der kontextgebundenen Verwendung) und betont damit, dass in der Evaluation die (oft vernachlässigte) Betrachtung von Kosten im weitesten Sinne (vgl. Checkpunkt 5) beachtet werden soll. *Merit* meint also den Wert eines Evaluationsgegenstandes ohne Berücksichtigung von Kosten jeder Art. *Worth* meint den Wert eines Evaluationsgegenstandes bei Berücksichtigung jedweder Kosten (monetäre und nicht-monetäre, auch externe Kosten) oder auch, der Wert eines Gegenstandes in Relation zu den aufzubringenden Ressourcen.
- 8 Wenn in diesem Text Personen angesprochen werden, die sowohl weiblichen wie männlichen Geschlechts sind, werden geschlechtsneutrale Bezeichnungen bevorzugt (z. B. Mitarbeitende) oder beide grammatikalischen Geschlechter genannt (Adressatinnen und Adressaten). Zwecks Erhöhung der Lesbarkeit werden Personenbezeichnungen in solchen Sätzen im Plural formuliert – auch wenn sie im engl. Original im Singular stehen –, selbstverständlich nur dann, wenn damit keine Sinnänderung einhergeht. Das generische Maskulinum wird gelegentlich zwecks Vereinfachung z. B. bei zusammengesetzten Wörtern oder bei feststehenden Begriffen genutzt (die grössten Experten unter den Experten).
- 9 Deutsche evaluationsspezifische Fachbegriffe können ausserdem im *Glossar wirkungsorientierter Evaluation* nachgelesen werden (<http://www.univation.org/glossar/index.php> [Stand 02.06.2006]).



HAUPTANMERKUNG C: Die Liste ist iterativ angelegt und nicht nur einmalig zu durchlaufen, d. h. sie ist wiederholt durchzugehen. Dies gilt auch für die Evaluationsplanung, da Klärungen, die sich in Folge späterer Punkte ergeben, oft Änderungen in Bezug auf vorhergehende Punkte erforderlich machen (und keine andere Anordnung der Checkpunkte kann dies vermeiden). Um mehr zum Charakter von Checklisten und ihrer Anwendung in der Evaluation zu erfahren, lesen Sie Scrivens Artikel zu diesem Thema¹⁰ sowie andere nützliche Beiträge über Checklisten in der Evaluation, die verschiedene Autorinnen und Autoren auf der Webseite „Projekt Checklisten“ veröffentlicht haben:

<http://evaluation.wmich.edu/checklists>

TEIL A: PRÄLIMINARIEN

Folgende Punkte I bis III sind grundlegende Bestandteile von Evaluationsberichten; dabei scheint es bloss so, dass sie keine grosse Relevanz für die Planungs- und Umsetzungsphasen der Evaluation haben. Es erweist sich hingegen als nützlich, die Situation des Berichtschreibens bereits am Anfang der Evaluation gedanklich durchzuspielen. Zum Beispiel kann es die Wichtigkeit der Erstellung eines begleitenden Logbuchs zur Evaluation verdeutlichen.

I. Management Summary

Ziel der „Management Summary“ in einem Evaluationsbericht ist, die Ergebnisse zusammenzufassen und *nicht* lediglich die Vorgehensweisen; üblicherweise werden die Vorgehensweisen sogar gar nicht erwähnt. (In anderen Worten, vermeiden Sie die schlechte Praxis, die Kurzfassung als „Aufreisser“ zu verwenden). Bedenken Sie während des gesamten Evaluationsprozesses, wie die Summary aussehen wird und wie sie sich auf die Bedarfe der Auftraggebenden und Stakeholder¹¹ sowie der Adressaten und Adressatinnen bezieht. Das wird Ihnen helfen, sich auf das Wichtigste zu konzentrieren. Die Management Summary sollte in der Regel eine selektive Zusammenfassung der Checkpunkte 11-15 darstellen und nicht mehr als eine bis maximal zwei Seiten umfassen, insbesondere wenn erwartet wird, dass Führungskräfte sie lesen.

II. Vorwort

Hier ist der Ort, sorgfältig Folgendes zu identifizieren und in Ihren Notizen festzuhalten: (i) den Auftraggebenden/die Auftraggebende (die Person, welche offiziell die Evaluation anfragt und – im Falle einer bezahlten Evaluation – für sie bezahlt oder die Bezahlung arrangiert und die — wie Sie hoffen — dieselbe Person ist, an die Sie berichten; falls dies nicht der Fall ist, versuchen Sie, dies aufzugleisen); (ii) die voraussichtlichen Adressaten und Adressatinnen (des Berichtes); (iii) die Stakeholder (die Personen mit beträchtlichem Interesse am Ergebnis der Evaluation) und (iv) alle anderen, welche die Evaluation verfolgen werden oder das Recht dazu haben, (a) die Evaluationsergebnisse und/oder (b) die Rohdaten zu sehen. Klären Sie Ihre Rolle — interner Evaluierender/interne Evaluierende, externer Evaluierender/externe Evaluierende, eine Mischung aus beiden (ein Außenstehender/eine Aussenstehende, der/die für kurze Zeit eingebracht wird, um das Personal zu unterstützen, einen Evaluationsprozess einzuleiten), ein Evaluationstrainer/eine Evaluationstrainerin (manchmal auch Empowerment-Evaluierender/-Evaluierende genannt) usw. Jede

10 Siehe: Scriven (2005)

11 Stakeholder sind direkt und indirekt Beteiligte und Betroffene des Evaluationsgegenstandes sowie alle anderen Personen oder Personengruppen mit einem Interesse an der Evaluation.



dieser Rollen ist mit verschiedenen Risiken und Erwartungen verbunden: seitens des Evaluationspersonals und Ihrer Kollegen und Kolleginnen, Auftraggebenden und dem Personal des evaluierten Programms usw.

Und jetzt ist es an der Zeit, zu den Details der Aufgabe/n zu kommen, so wie die Auftraggebenden sie sehen — oder ermutigen Sie die Auftraggebenden, Details zu klären, an die sie bisher nicht gedacht haben. Können Sie den Hintergrund und die Art der Anfrage, den zugrunde liegenden Bedarf oder das Interesse, das zur Evaluation führt, bestimmen? Zum Beispiel: Besteht die Anfrage oder der Bedarf eher für eine Evaluation von kostengebundenem Wert¹² — was ernsthafte Kostenanalyse einschliesst — oder für die Evaluation von kostenfreiem Wert (oder von Bedeutung) oder mehr als eins von diesen dreien? Was genau sollen Sie evaluieren: Wie viel Kontext sollte einbezogen werden? Wie viele Details sind wichtig? Sind die Zwecksetzung, die Effekte des gesamten Programms oder die einzelner Komponenten des Programms oder vielleicht zusätzlich die Programmtheorie der Auftraggebenden bezüglich der Funktion einzelner Komponenten zu evaluieren? Sollen Sie Globalauswirkungen «*impacts*»¹³ in jeglicher relevanten oder in ausgewählter Hinsicht berücksichtigen? Ist die Evaluation formativ, summativ, askriptiv¹⁴ oder soll sie mehreres davon leisten? Soll die Evaluation Benotung «*grading*», Rangordnung «*ranking*», Punktbewertung «*scores*», Profile «*profiles*» oder Zuteilungsregeln «*apportionments*» erbringen?¹⁵ Sind Empfehlungen, Erklärungen (d. h. Ihre eigene Theorie), Fehlerdiagnosen oder Vorhersagen gefragt, erwartet oder durchführbar? Sind die Auftraggebenden *wirklich* bereit und motiviert, aus Fehlern zu lernen oder ist das nur eine Pose? (Ihr Vertrag oder, falls Sie intern evaluieren, Ihr Job hängt von der richtigen Antwort auf diese Frage ab. Sie können den folgenden Test erwägen: Fragen Sie die Auftraggebenden, wie mit der Entdeckung gravierend ernster Fehler umgegangen würde. — Die Reaktion wird Ihnen eine Idee geben, ob der/die Auftraggebende das Zeug zum/zur „guten Auftraggebenden“ hat.) Wurde Unterstützung zur Interpretation und Nutzbarmachung nach der Berichterstattung angefragt? (Wenn nicht, bieten Sie dies zuschlagsfrei an — siehe Checkpunkt 12.)

ANMERKUNG: Am besten ist es, sich erst nach einem schnellen Testlauf durch die KEC zu verpflichten und en detail zu diskutieren, was durchführbar zu evaluieren ist. Also bitten Sie um etwas Zeit, um dies zu tun, über Nacht wenn möglich. Verschriftlichen Sie jegliche verhandelten Änderungen gegenüber dem eingangs Ausgehandelten. An dieser Stelle formulieren Sie auch Anerkennungen/Danksagungen...

III. Methodologie

Jetzt, wo die Fragestellungen klar sind, wie werden Sie die Antworten finden? Beispiele von reflexiven Fragen, die hier geklärt werden müssen: Haben Sie angemessene Expertise im jeweiligen Feld? Falls nicht, wie werden Sie das Evaluationsteam mit dieser ergänzen (durch Berater und Beraterinnen, ein Beratungsgremium, durch ein volles Teammitglied)? Können Sie Kontroll- oder Vergleichsgruppen verwenden,

12 Siehe Fussnote 7.

13 Auswirkungen des Evaluationsgegenstandes, die über direkte intendierte Einflüsse auf Programmteilnehmende hinausgehen und nicht nur deren Wohlergehen betreffen, sondern auch das anderer, (a) die nicht am Programm teilgenommen haben oder (b) die mit den Zielgruppen verbunden sind (z. B. ihre Familienangehörigen) oder (c) die nicht zur direkten Zielgruppe gehören (z. B. Programmpersonal), oder (d) Auswirkungen auf grössere soziale Einheiten wie Organisationen, Kommunen oder die Gesellschaft.

14 Scrivens neueste Unterscheidung, im Sinne von Evaluation der Evaluation wegen «*for the sake of it*».

15 Siehe: *Scriven* (1998)



um die Kausalität angenommener Wirkungen zu bestimmen? Falls eine Kontrollgruppe vorhanden ist, können Sie Personen nach dem Zufallsprinzip zuteilen? Kann die Studie doppel- oder einfachblind durchgeführt werden? Wenn eine Stichprobe verwendet wird, wie wurden die Elemente ausgewählt und falls stratifiziert, wie wurde geschichtet? Falls keine dieser Ansätze möglich ist: Wie werden Sie die Kausalität (von Wirkungen des Evaluationsgegenstands) bestimmen (je nach Aufgabenstellung kann es erforderlich sein, dass Sie den relativen Beitrag verschiedener Komponenten des Evaluationsgegenstandes zu den Wirkungen bestimmen müssen)? Ist oder sollte die Evaluation zielgeführt oder zielfrei angelegt sein? Falls Einschätzungen durch Beurteilende «judges» vorgenommen werden, welche Massnahmen zur Kontrolle von Zuverlässigkeit und Verzerrungen sind notwendig (um Glaubwürdigkeit und Validität zu garantieren)? Wie werden Sie nach Nebenfolgen suchen? Legen Sie so früh wie möglich in der Evaluation verwendete Verfahren zur Datensammlung (wie z. B. nicht-teilnehmende oder teilnehmende Beobachtungen, Protokollieren, Journal-Schreiben, Audio-/Video- Aufnahmen, Tests, Simulationen, Rollenspiele, Umfragen, Interviews, Fokusgruppen, Textanalyse, Recherche in Bibliotheken, online und durch Suchmaschinen usw.) und Datenanalyse (wie z. B. Statistik, Kostenanalyse, Modellierung, Expertenberatung usw.) fest sowie Berichterstattungsmethoden (Text, Geschichten, Theaterstücke, grafische Darstellungen, Freistilzeichnungen, Fotos, Filme usw.) und die Begründung für ihren Einsatz (was eine Literaturlauswertung zu einigen dieser Methoden verlangen mag). Zusammengefasst, legen Sie die “Logik der Evaluation” dar, d. h. die umfassende Begründung des Evaluationsplans.

TEIL B: GRUNDLAGEN

In der folgenden Zusammenstellung von Untersuchungen werden der Kontext und die Natur des Programms dargelegt, um die Arbeit an den Schlüsseldimensionen des Wertes des Evaluationsgegenstandes im Abschnitt C beginnen zu können.

1. Hintergrund und Kontext

Identifizieren Sie das historische, jüngste, aktuelle und künftige Umfeld des Programms. Identifizieren Sie: (i) alle „vorgeschalteten Stakeholder“ «*upstream stakeholders*» — und ihre Einsätze «*stakes*» — nicht nur der Auftraggebenden (d. h. Personen, Gruppen oder Organisationen, die bei der Implementierung des Programms oder seiner Evaluation assistiert haben, z. B. durch Finanzierung, Beratung oder die Bereitstellung von Räumen); (ii) ermöglichende (und in jüngster Zeit relevante) Gesetzgebung/Politik und jegliche Änderungen in der Gesetzgebung/Exekutive seit Einführung des Programms; (iii) das zugrunde liegende Konzept, auch offizielle Programmtheorie oder die politische Logik (falls eines der beiden existiert oder zuverlässig erschlossen werden kann; obwohl beide nicht notwendig sind, so sind sie doch manchmal zweckdienlich); (iv) allgemeine Ergebnisse der Literaturlauswertung zu ähnlichen Interventionen (einschliesslich grauer Studien, die nicht in den handelsüblichen Medien veröffentlicht sind) und des Internets (berücksichtigen Sie das „Invisible Net“¹⁶ sowie die neuesten Newsgroup- und Blogsuchmaschinen); (v) vorherige Evaluationen, falls vorhanden; (vi) ihre Globalauswirkungen, falls vorhanden.

2. Beschreibungen und Definitionen

Protokollieren Sie jegliche offizielle Beschreibung des Programms + seiner Komponenten + seines Kontexts/Umfeldes, aber unterstellen Sie nicht ungeprüft, dass diese Beschreibungen korrekt sind. Entwickeln Sie eine eigene korrekte und vollständige, genügend detaillierte Beschreibung (diese könnte sehr anders sein), um den Evalua-

16 Von Suchmaschinen noch nicht indexierte oder nicht frei zugängliche Teile des Internet.

tionsgegenstand zu identifizieren und vielleicht — abhängig vom Zweck der Evaluation — um ihn replizieren zu können. Erstellen Sie eine detaillierte Beschreibung der Ziele/Meilensteine (falls Sie nicht im zielfreien Modus agieren). Erläutern Sie die Bedeutung jeglicher Fachbegriffe, speziell solcher, die der voraussichtlichen Adressatengruppe nicht bekannt sind. Vermerken Sie bedeutende Muster/Analogien/Metaphern, die von Teilnehmenden verwendet werden (oder die implizit sind) oder welche Ihnen bekannt werden. Dies sind potenzielle Beschreibungen, die sich später als klärender herausstellen können als wortgetreue Wiedergabe von Selbstbeschreibungen, ob sie gerechtfertigt sind oder nicht. Unterscheiden Sie die Bemühungen der Initiatoren oder der Initiatorinnen, ein Programm ins Leben zu rufen vom Programm selbst; beides sind Interventionsbündel, doch nur die letztere stellt (normalerweise) den Evaluationsgegenstand dar.

3. Konsumierende «*impactees*»¹⁷

Konsumierende umfassen (i) Empfangende/Nutzende der Dienstleistungen/Produkte (d. h. die nachgeschalteten «*downstream*» direkten *impactees*), manchmal auch „Klienten oder Klientinnen“ genannt (jedoch sind sie Klienten oder Klientinnen des Programmes und nicht die Auftraggebenden¹⁸ der Evaluation; so ist es meist ratsam, die Verwendung des Begriffs im Kontext der Evaluation ausschliesslich für Geldgebende zu verwenden), PLUS (ii) nachgeschaltete indirekte *impactees* (z. B. Familien oder Mitarbeitende der Empfangenden, die durch Nachwirkungen «*via ripple effect*» beeinflusst werden). Mitarbeitende des Programms sind auch *impactees*, doch wir behandeln sie separat (und nennen sie zwischengeschaltet «*midstream*»), weil die Verpflichtungen ihnen gegenüber in den meisten Ansätzen der Programmevaluation ganz andere und viel schwächer ausgeprägte sind (ihr Wohlergehen ist nicht die Existenzberechtigung für das Programm). Geldgebende, Steuerzahlende und politische Unterstützer und Unterstützerinnen sind gewissermassen auch *impactees* und werden ebenfalls anders behandelt (und vorgeschaltete «*upstream*» *impactees* oder manchmal Stakeholder genannt, obwohl der Begriff Stakeholder häufig weiter gefasst wird und alle *impactees* einschliesst), es sei denn, sie sind gleichzeitig direkt Empfänger oder Empfängerinnen «*recipients*». Beachten Sie, dass es weitere vorgeschaltete *impactees* gibt, die weder Geldgebende noch Empfänger/Empfängerinnen der Dienstleistungen/Produkte sind, jedoch auf die Ankündigung oder Planung des Programms reagieren, bevor es realisiert wird (wir nennen sie Vorwegnehmende «*anticipators*»). Wenn Sie Konsumierende identifizieren, seien Sie darauf gefasst, dass diese häufig weder den Namen noch die Ziele des Programms kennen und vielleicht nicht wissen, dass sie vom Programm beeinflusst wurden oder dass das Programm an sie gerichtet war. (Eventuell müssen Sie die Tracer- und/oder Modus

17 *Impactees* sind alle vom Evaluationsgegenstand beeinflussten Personen – sowohl die direkt und gezielt Beeinflussten (dann treten bei ihnen Outcomes auf) als auch die nicht-intendiert und indirekt Beeinflussten (i. S. von Globalauswirkungen – «*impacts*», wie Angehörige in Familien, andere Mitglieder in Organisationen oder sonstige Bewohner und Bewohnerinnen von Kommunen. Mehr zum Thema *consumer/impactee* bei Davidson (2005).

18 In der englischsprachigen Evaluationsliteratur werden Auftraggebende entweder als *sponsors* bezeichnet, was ihre Finanzierungsrolle betont, oder als *clients*; dort, noch mehr als im deutschsprachigen Raum sind mit „Klienten und Klientinnen“ meist die Zielgruppen gemeint, die Scriven mit „*consumers*“ oder „*impactees*“ anspricht. Scriven meint den Terminus „Zielgruppen“, da durch ein Programm auch Personen erreicht werden, die nicht zur Zielgruppe gehören, neben den von ihm genannten nach- und zwischengeschalteten *impactees* können dies auch nicht-intendierte Nutzende sein.

Operandi-Methodologie¹⁹ einsetzen.) Während Sie nach der beeinflussten Population suchen, können Sie auch einbeziehen, welche anderen Gruppen durch Variationen des Programms hätten beeinflusst oder vor dem Einfluss des Programms geschützt werden können: Diese Gruppen definieren potenzielle alternative Einflusspopulationen, welche Möglichkeiten der Ausweitung oder der Einschränkung des Programms andeuten könnten, wenn/falls Sie zum Checkpunkt 12 gelangen.

4. Ressourcen (auch Stärkenabschätzung «*strengths assessment*»)

Finanzielle, physische und intellektuell-sozial-relationale Ressourcen des Programms. Diese beinhalten Fähigkeiten, Wissen und Motivation des Personals, von Freiwilligen, Mitgliedern der Gemeinde und von anderen Unterstützenden. Dies sollte umfassen, was aktuell oder in der Vergangenheit hätte verwendet werden *können*, und nicht nur das, was tatsächlich verwendet *wurde*. Dies definiert den „Möglichkeitsraum“, d. h. die Reichweite dessen, was hätte getan werden können, was oft ein wichtiges Element für Vergleiche ist, welche die Evaluation anstellt. Es könnte auch hilfreich sein, spezifische Ressourcen aufzulisten, die bei der Umsetzung nicht verwendet wurden oder nicht zur Verfügung standen. Zum Beispiel: in welchem Masse wurden potenzielle *impactees*, Stakeholder, Spendensammelnde, Freiwillige und mögliche Spender/Spenderinnen *nicht* rekrutiert oder *nicht* in dem Masse einbezogen, in dem sie hätten einbezogen werden können? Zur Überprüfung und Ergänzung ziehen Sie alle einengenden Faktoren «*constraints*» des Programms in Betracht.

5. Werte²⁰

Beginnen Sie mit Hilfe untenstehender Liste, allgemein relevante Werte (d. h. Gütekriterien und -dimensionen) für den Evaluationsgegenstand mit seinen besonderen Randbedingungen zu identifizieren. Validieren Sie die Werte für das derzeitige Evaluationsprojekt, wie aktuell und kontextbezogen angemessen. Danach setzen Sie präzise Standards (die spezifischen Werte) für diese Evaluation. Das bedeutet erstens, eine Gradeinteilung «*grading scale*» für jede Dimension zuzuweisen, z. B. mag für einen standardisierten Mathematiktest ein Ergebnis «*score*» von 65% gefordert werden. Ergänzen Sie dies um die vorläufige Angabe von „Gewichtungen“ «*stars*», „Mindestanforderungen“ «*bars*» und „Stufen“ «*steps*» soweit auf den Evaluationsgegenstand in dieser (oder diesen) Implementierung(en) anwendbar. Die Gewichtungen sind Bedeutsamkeitsgewichte, d. h. die relative oder absolute Relevanz der verschiedenen Gütedimensionen (oder anderen Werte), die herangezogen werden, um von den Fakten über den Evaluationsgegenstand, die Sie lokalisieren oder bestimmen, zu den evaluativen Schlussfolgerungen zu gelangen. (Die Relevanz/Wichtigkeit kann qualitativ [z. B. durch Kategorien wie wesentlich, mittelgross, unwesentlich], quantitativ [z. B. durch Punkte auf einer 5- oder 10-Punktskala oder — normalerweise eine bessere Methode — durch die Verteilung von 100 „Gewichtspunkten“ über das Gesamtset der Werte] oder relativ [durch die Rangordnung der Werte nach ihrer Wichtigkeit] ausgedrückt werden.) Die Mindestanforderungen «*bars*» sind *absolute* Minimalstandards für die Akzeptanz des Evaluationsgegenstands, falls vorhanden. Das heisst, sie sind Minima auf einer bestimmten Skala, die erreicht werden müssen, soll der ‚Kandidat‘ akzeptiert werden, *ungeachtet dessen wie gut er auf anderen Skalen punktet*. Beachten Sie, dass eine ungenügende Zensur für die Leistung auf einer bestimmten Skala nicht automatisch „Misserfolg hinsichtlich der Mindestanforderun-

19 Ableitung der Ursachen für Wirkungen, indem man Ereignisse, Prozesse oder Eigenschaften untersucht, die mit den Effekten zu tun haben; analog zu Verfahrensweisen in der Kriminalistik.

20 Dieser Checkpunkt wurde uns am 31. März 2006 von Michael Scriven für die Übersetzung in überarbeiteter Form zur Verfügung gestellt.

gen“ bedeutet. Zum Beispiel kann eine ungenügende Zensur auf der Mathematikskala akzeptabel sein, wenn sie durch andere Stärken *für die Zulassung in ein Programm für kreatives Schreiben* ausgeglichen wird. Mindestanforderungen und Gewichtungen können festgesetzt werden für jegliche relevante Eigenschaften (d. h. Gütedimensionen) oder für Dimensionen gemessener (oder gewerteter) Leistungen und können ausserdem ganzheitliche Gewichtungen und Mindestanforderungen enthalten.²¹ Die Grade, Gewichte und Mindestanforderungen werden häufig Standards genannt. In seriöser Evaluation kann es zudem angemessen sein, Stufen zu setzen, d. h. Punkte oder Intervalle auf gemessenen Gütedimensionen, bei denen sich die Gewichtung verändert, wenn die Gradänderung nicht genügend Wirkung zeigt. (Mindestanforderungen und Stufen können sowohl unscharf als auch präzise sein.)

Zumindest die folgenden Werte sollten auf ihre Relevanz für die Evaluation erwogen werden; suchen Sie ausserdem nach weiteren:

- (i) Bedürfnisse der beeinflussten Population durch Bedarfsermittlung (unterscheiden Sie Leistungs-Bedürfnisse «*performance needs*» von Behandlungs-Bedürfnissen «*treatment needs*», befriedigte von nicht befriedigten Bedürfnissen und Bedürfnisse, die befriedigt werden können, von solchen, die prinzipiell befriedigbar aber mit den gegebenen Ressourcen unmöglich zu befriedigenden sind [berücksichtige den Checkpunkt 4 Ressourcen]). Bedürfnisse sind Tatsachen, nicht Werte an sich, doch in allen Kontexten, die zumindest rudimentäre ethische Erwägungen zulassen, sind diese Fakten von Werten durchdrungen.
- (ii) Gütekriterien die mit der Definition des Evaluationsgegenstandes verbunden sind sowie aus der Anwendung von Standards resultieren. (z. B. Da Programme häufig per definitionem dann als besser angesehen werden, wenn mehr Personen erreicht werden oder wenn [unter gleichen Bedingungen] grössere positive Effekte bei diesen erzielt werden, schliessen Gütekriterien normalerweise die Anzahl der vom Programm beeinflussten Personen und die gewünschte Wirkungstiefe ein.)
- (iii) Logische Anforderungen (z. B. Konsistenz)
- (iv) Legale und (v) ethische Anforderungen (sie überschneiden sich) und schliessen normalerweise (angemessene) Sicherheit, Vertraulichkeit und vielleicht Anonymität für alle *impactees* ein.
- (vi) Persönliche und organisatorische Ziele/Wünsche, falls sie nicht in Konflikt mit ethischen/legalen/praktischen Erwägungen stehen (und soweit Sie nicht eine zielfreie Evaluation durchführen); diese sind normalerweise weniger relevant als die Bedürfnisse der *impactees*, doch sie sind geeignet, evaluative Schlussfolgerungen anzuregen. Ein Beispiel dafür aus dem Alltagsleben ist die Fragestellung danach, welche Wohnung man mieten soll und die damit verbundenen persönlichen Präferenzen.
- (vii) Originalgetreues Aufgreifen von behaupteten Spezifizierungen („Authentizität“, „Einhaltung“ «*adherence*» oder „Befolgung“ «*compliance*» — dies ist häufig

21 **Fussnote 1 aus der KEC** – Ein Beispiel: Die Bewerber und Bewerberinnen für ein Programm für Studierende mit abgeschlossenem Studium (z. B. Doktorierende) erfüllen alle dimension-spezifischen Minimalstandards (d. h. sie halten die Mindestanforderungen), doch sie erreichen die Mindestvoraussetzungen in so vielen Dimensionen so knapp (und sie zeigen Schwächen in Dimensionen, für die keine Mindestanforderungen festgelegt sind), dass das Auswahlkomitee denkt, dass sie für das Programm nicht genügend geeignet sind. Wir können dies als einen Fall beschreiben, in dem die ganzheitlichen Mindestanforderungen nicht getroffen wurden, was im Beispiel implizit ist, aber oft erst durch einen Dialog explizit gemacht werden kann.

- sinnvoll durch einen „Implementierungsindex“ ausdrückbar); und —ähnlich — „Konsistenz“ mit dem unterstellten Programmmodell (falls dieses Modell ausserhalb eines akzeptablen Zweifels steht).
- (viii) Unterhalb der Gesetzesschwelle liegende, jedoch in Bezug auf mögliche Gesetzgebung wichtige Präferenzen
 - (ix) Professionelle Qualitätsstandards, welche auf den Evaluationsgegenstand zutreffen^{22,23}
 - (x) Expertenurteile
 - (xi) Historische/traditionelle/kulturelle Standards
 - (xii) Wissenschaftlicher (kostenfreier) Wert (oder kostenabhängiger Wert oder Bedeutsamkeit)
 - (xiii) Technologischer (kostenfreier) Wert (oder kostenabhängiger Wert oder Bedeutsamkeit)
 - (xiv) Marktfähigkeit
 - (xv) Politischer (kostenfreier) Wert, falls Sie diesen jenseits akzeptablen Zweifels bestimmen können
 - (xvi) Zuletzt, aber an sich nicht am wenigsten bedeutend — Ressourcen-Wirtschaftlichkeit (d. h. wie wenig wirksam «*low-impact*» ist das Programm in Relation zu Geld, Raum, Zeit, Arbeit, Kontakten, Expertise und Ökosystem?)

Natürlich ist die Identifizierung/Validierung/Anwendung mancher dieser Werte in manchen Fällen unwichtig und in anderen von entscheidender Bedeutung; manchmal ist sie leicht und manchmal sehr kompliziert; häufig ist Rat von Experten/Expertinnen und/oder von *impactees*/Stakeholdern erforderlich. Ausserdem werden einige natürlich mit anderen in Konflikt stehen (z. B. Wirkungsgrösse mit Ökonomie), so dass deren relatives Gewicht für diesen besonderen Fall bestimmt werden muss, eine nicht triviale Aufgabe. Daher sollten Sie nicht selbstverständlich voraussetzen, dass eine Rangordnung «*ranking*» aus der Evaluation entstehen muss. Falls das nicht notwendig oder nützlich ist, können Sie häufig die relative Gewichtungen unterlassen oder zumindest der Anforderung auf Präzision bei der Erstellung entgehen, indem Sie den Evaluationsgegenstand einfach benoten oder ein Profil (Darstellung der Güte auf allen relevanten Gütedimensionen in Form von Balkendiagrammen) erstellen.

ANMERKUNG: Sie müssen in diesem Checkpunkt *alle* Werte einschliessen, die Sie für die Evaluation verwenden wollen, auch solche die für die Evaluation von *Nebenresultaten* (falls vorhanden) verwendet werden, nicht nur die für die *intendierten* Effekte (falls vorhanden). Manche dieser Werte werden Ihnen erst dann auffallen, wenn Sie die Nebenfolgen bestimmen (Checkpunkt 7), doch das ist kein Problem — diese Checkliste ist iterativ, was bedeutet, dass Sie oft zurückgehen werden, um Erkenntnisse zu früheren Checkpunkten zu modifizieren.

-
- 22 **Fussnote 2 aus der KEC:** Da in der Metaevaluation die Evaluation selbst der Evaluationsgegenstand ist, werden auch professionelle Evaluationsstandards eingeschlossen, wenn Sie zu diesem Checkpunkt gelangen. Die besten derzeit verfügbaren Standards, sind die vom *Joint Committee* entwickelten (*The Program Evaluation Standards, 2nd ed.*, Sage (1994); deutsch: *Sanders* (2006)).
 - 23 Für den deutschsprachigen Kontext, siehe auch die DeGEval- und SEVAL-Standards (Nachweis im Literaturverzeichnis).

TEIL C: UNTER-EVALUATIONEN

Jede dieser Unter-Evaluationen umfasst (i) eine Faktfindungsphase gefolgt von (ii) einem Prozess, in welchem die Fakten mit den ihnen zugrunde liegenden Werten (von Checkpunkt 5) kombiniert werden, was die Unter-Evaluation hervorbringt. In anderen Worten, im Teil C werden fünf separate Schritte vom „Was so?“ zum „So was?“ erfordert, z. B. von den als „XXX gemessenen Effekten“ zu den „im höchsten Maße nützlichen Effekten“ (oder dem „Schnäppchen“ etc.).

6. Prozess

Dies ist die Einschätzung «*assessment*» von kostenfreiem und kostenabhängigem Wert sowie von Bedeutung, von allem was im Programm passiert, bevor wirkliche Outcomes entstehen; dabei handelt es sich insbesondere um Vision, Konzeption, Planung und Durchführung des Programms, beginnend bei der Rechtfertigung der Ziele (falls Sie nicht im zielfreien Modus handeln), welche sich geändert haben können oder sich gerade ändern über Vorkehrungen in der Konzeption, um sie gegenüber Umwelt-, politischen oder finanziellen Zwängen belastbar zu machen (einschliesslich der Planungen für den Worst Case); Implementierungstreue (d. h. der Grad der Implementierung des angenommenen Programms in seiner Urform, im Sinne von „Authentizität“, „Einhaltung“ oder „Befolgung“); Exaktheit seines offiziellen Namens (falls dieser beschreibend ist), seines Untertitels oder der Programmbeschreibung (z. B. „ein forschungsbasiertes naturwissenschaftliches Bildungsprogramm für Mittelschulen“, „den Leistungsstand anheben“, „ein Trainingsprogramm für kritisches Denken“); Programmsteuerung; Aktivitäten; Vorgehensweisen; der Lernprozess; Einstellungen/Werte; Moral; vielleicht auch, falls Sie dies einschliessen, die Qualität der originären Programmlogik und der aktuellen Logik (beide, die aktuelle offizielle und die möglicherweise sich davon unterscheidende, welche implizit in den Programmoperationen/dem Handeln des Programmpersonals enthalten ist). Prozessevaluation kann auch die Evaluation von dem umfassen, was häufig „Outputs“ genannt wird (üblicherweise als „intermediäre Outcomes“ auf dem Weg zu wirklichen Outcomes auf dem Weg zu Globalresultaten), wie z. B. Änderungen in Wissen, Fertigkeiten und Einstellungen beim Personal (oder bei Auftraggebenden), wenn diese Veränderungen nicht wesentliche Outcomes mit eigenem Recht sind.

7. Outcomes

Evaluation von (positiven und negativen) Effekten auf Konsumierende: direkte/indirekte; beabsichtigte/unbeabsichtigte; unmittelbare/kurzfristige/langfristige. Outcomes können im Rahmen einer Methodologie des Hypothesentestens nicht identifiziert werden, weil die wichtigsten Effekte oft unvorausehbare sind. (Die beiden grundlegenden Methoden zur Identifikation von Nebenresultaten sind die zielfreie Evaluation und das legendäre „Buch der Ursachen“ «*Book of Causes*»²⁴. Unmittelbare Auswirkungen werden oft Outputs genannt, insbesondere wenn ihre Rolle die ei-

24 **Fussnote 3 aus der KEC:** Wenn das „Buch der Ursachen“ für einen Namen eines Faktors oder eines Ereignisses aufgeschlagen wird, zeigt es (i) auf der linken Seite (verso) in manchen Fällen die als verursachend bekannten Dinge und (ii) auf der rechten Seite (recto) alle Dinge, die es verursachen kann: Das ist die Seite, die Sie benötigen, damit Ihre Suche nach Nebenresultaten erfolgreich verläuft. Da das „Buch der Ursachen“ leider nur ein virtuelles Buch ist, müssen Sie diese Seiten selbst kreieren, indem Sie alle Ihnen zur Verfügung stehenden Ressourcen verwenden wie z. B. zugängliche Expertise und Literatur-/Internetrecherchen. Gute forensische Pathologen und gute Epidemiologen im Feld, neben anderen Wissenschaftlern und Wissenschaftlerinnen, haben sehr umfassende „lokale Ausgaben“ des Buchs der Ursachen im Kopf.

ner zeitlich vorgelagerten oder intendierten Ursache von Haupt-Outcomes ist; sie sind normalerweise durch Checkpunkt 6 abgedeckt. Doch beachten Sie, dass einige wirkliche Outcomes (d. h. Resultate, welche von wesentlicher Bedeutung sind, seien sie beabsichtigt oder nicht) während des Prozesses anfallen und hier zu berücksichtigen sind. (Langzeitresultate werden manchmal Effekte genannt [oder „wahre Effekte“ oder „Resultate“] und die Gesamtheit dieser wird häufig Impact genannt; Sie können die Verwendung dieser Begriffe den Präferenzen der Auftraggebenden/Adressaten und Adressatinnen/Stakeholder anpassen.) Manchmal, doch nicht immer, ist es nützlich und möglich, Erklärungen zu Erfolgen/Misserfolgen hinsichtlich Komponenten/Kontext/Entscheidungen anzubieten. Dies kann, muss aber nicht, die Identifizierung der tatsächlich ablaufenden Programmlogik/-theorie erfordern (im Kontrast zu den [i] ursprünglichen, [ii] aktuellen, [iii] offiziellen und [iv] impliziten Logiken oder Theorien). (Siehe Checkpunkt 12) Vorausgesetzt, dass die wichtigsten Outcomes unbeabsichtigt, sogar unvorhersehbar sein können, lohnt es sich, zwischen *Nebeneffekten* (die die Zielgruppe beeinflussen) und *Nebenimpacts* (im Sinne von Einflüssen auf Nicht-Zielgruppen) zu unterscheiden. Bedenken Sie, dass als Erfolg zu bezeichnende Fälle ungeachtet von durchschnittlichen Verbesserungen eine individuelle Behandlung erfordern können (da der Nutzen für sich allein die Kosten des Programms zu rechtfertigen vermag); falls dies der Fall ist, kann dies auch auf Misserfolgswfälle zutreffen. Verfolgen Sie den Ansatz der „dreifachen Basis“ «*triple bottom line*», d. h., suchen Sie nach (ii) sozialen und (iii) Umwelt-Einflüssen, sowie nach (i) konventionellen Outcomes. Schliesslich vergessen Sie nicht (i) Effekte auf das Programmpersonal, insbesondere gelernte Lektionen und erworbene Fähigkeiten, und (ii) die vor dem Programm entstehenden Effekte: Das sind (häufig die grössten) Effekte, die sich aus der blossen Ankündigung des Programms oder der Entdeckung ergeben, dass ein Programm implementiert *wird* oder auch nur *vielleicht* implementiert wird. Diese Effekte schliessen Booms im Immobilienhandel, bei der Einwanderung zur und Abwanderung von der Gemeinde ein und sind oft ernster (zumindest in der ökonomischen Dimension) als die direkt durch die Programmimplementierung verursachten Resultate.

8. Kosten

Erfassen Sie, sofern relevant und möglich, monetäre *und* nicht-monetäre Kosten, direkte *und* indirekte Kosten, pagatorische *und* Opportunitätskosten; gliedern Sie diese nach Entwicklungsphasen (d. h. Kosten für Erst-Inbetriebnahme, Wartung, Erweiterung und Schliessung) und/oder nach der Kalenderperiode und/oder nach Kostenarten (Miete, Ausstattung, Personal etc.). Rechnen Sie, falls vorhanden, aufgewendete, doch niemals realisierte Werte ein, z. B. soziales Kapital. Die häufigsten nicht-monetären Kosten sind Räume, Zeit, Expertise, Arbeitsaufwand (falls diese nicht über den Markt zugänglich sind (ansonsten werden sie zu geldlichen Kosten) und Stress, politisches und persönliches Kapital sowie Beeinträchtigungen der Umwelt, welche selten *völlig* durch Geld abgedeckt werden können.

9. Vergleiche

Üblicherweise werden Vergleiche angestellt gegenüber anderen Mitteln, mit denen der gleiche oder ähnlicher Nutzen bei etwa gleichen oder weniger Ressourcen erreicht werden könnte; was dem standhalten kann ist als kritischer Wettbewerber «*critical competitor*» anzusehen. Normalerweise lohnt es sich, erstens nach einer Alternative zu suchen, die deutlich billiger, aber fast genauso effektiv ist wie der Evaluationsgegenstand (= *el cheapo*), und zweitens nach einer viel stärkeren dabei auch kostspieligeren Alternative, mit deutlich grösseren Erträgen «*payoffs*» und Prozessvorteilen (= *el magnifico*); und manchmal lohnt es sich auch, den Evaluationsgegenstand mit einer breit angewendeten und/oder anerkannten alternativen Herangehens-



weise zu vergleichen, obwohl diese nicht wirklich mit im Rennen ist (z. B. eine „lokale Ikone“). Bedenken Sie, dass „den gleichen Effekt zu haben“ sowohl Nebeneffekte wie auch intendierte Effekte einschliesst. Wenn dies auch auf dünnem Eise erfolgt, gibt es doch manchmal gute Gründe, einen Vergleich mit einer als möglich demonstrierbaren Alternative anzustellen (der nächste Checkpunkt gibt für einen solchen Fall Anregungen). Das Eis ist dünn, weil Sie jetzt aus der Rolle des/der Evaluierenden in die des/der Programmgestaltenden «*program designer*» wechseln, was zu einem Interessenkonflikt führen kann, falls Sie eine fortlaufende Aufgabe in der formativen Evaluation haben — und Sie müssen sich sicher sein, dass die Einladung an Sie, dies zu tun, echt und entschieden war (siehe auch Checkpunkt 12).

10. Generalisierbarkeit

(Auch Exportierbarkeit, Übertragbarkeit, Transportierbarkeit genannt; ungefähr das gleiche wie *Campbells* „externe Validität“, doch unter Generalisierbarkeit werden auch Nachhaltigkeit, Haltbarkeit/Dauerhaftigkeit und Belastbarkeit verstanden.)

Ergibt dieses Programm ähnliche Resultate, wenn es mit anderen Inhalten stattfindet, an einem anderen Ort, mit anderem Personal, mit anderen Empfängern und Empfängerinnen in einem anderen (sozialen, politischen, physikalischen) Klima, usw.? Generalisierbarkeit zu späteren Zeitpunkten meint Langlebigkeit (und unter widrigen Bedingungen ist es Haltbarkeit) und es ist fast immer kritisch, dies abzuwägen. Wenn Sie sicher stellen, dass dieser Checkpunkt finanzielle, soziale, räumliche, zeitliche, umgebungsbedingte, politische und andere nicht-geldliche Kosten sowie Kapazität und Überlebensbedingungen umfasst, wird eine Nachhaltigkeitsbewertung möglich (auch Risikobelastbarkeit «*resilience to risk*» genannt), was noch wichtiger ist als Langlebigkeit, zum Beispiel wenn internationale oder interkulturelle Entwicklungsprogramme evaluiert werden.

ANMERKUNG: Was Sie dann generalisieren, ist das „Programm im Kontext“ und der Kontext sollte vorzugsweise näher beschrieben sein und die Infrastruktur beinhalten. Hier, wie auch im letzten Satz in Checkpunkt 9, sprechen wir von Möglichkeiten, welche – obwohl riskant – manchmal die grössten Beiträge zur Evaluation selbst und zur Besserung der Welt erzeugen. Sehen Sie auch „mögliche Szenarien“ unter Checkpunkt 14. (Alle drei zeigen, dass bedeutsame Evaluation ein kreatives und nicht bloss ein reaktives Unternehmen ist.)

TEIL D: SCHLUSSFOLGERUNGEN

11. Gesamtbedeutung

Fügen Sie die Unter-Evaluationen aus Teil C zu einer Gesamtevaluation zusammen, d. h. zumindest in ein Profil (ein Mittel, um eine multidimensionale Schlussfolgerung darzustellen) oder in eine eindimensionale Schlussfolgerung — eine Note oder einen Rangplatz, falls dies verlangt wird (normalerweise ist das viel schwieriger). Der Fokus (die Perspektive) sollte normalerweise der gegenwärtige und zukünftige Einfluss auf die Bedürfnisse der Konsumierenden sein und ethischen und gesetzlichen Vorgaben entsprechen (sowie der Durchführbarkeit etc., d. h. alle anderen relevanten Werte einschliessend). Doch normalerweise sollten auch einige Schlussfolgerungen vorhanden sein, die sich auf die Informationsbedürfnisse (und Wünsche und Hoffnungen, falls anwendbar) der Auftraggebenden (und anderer relevanter Stakeholder) beziehen, z. B. erreichte Ziele oder unrealisierter Wert, falls kalkulierbar.

12. Fakultative Empfehlungen und Erklärungen

Das allgemeine Prinzip betreffend Empfehlungen und Erklärungen kann dank *Gloria Steinem*²⁵ wie folgt ausgedrückt werden: Eine Evaluation ohne Empfehlungen (oder Erklärungen) ist „wie ein Fisch ohne Fahrrad“. Dennoch bestehen dort einige Fallstricke. Mikro-Empfehlungen — solche, die *interne* Funktionsweisen des Programmmanagements und die Wahl oder Verwendung von Ausstattung/Ausrüstung betreffen — werden für die Evaluierenden im Lauf der Untersuchungen häufig offensichtlich und sind unter geringen oder gar ohne zusätzliche Kosten/Aufwendungen leistbar (man sagt, sie „fallen ab“ bei der Evaluation) *oder* sie mögen den gewieften Evaluierenden, die dem Programm helfen möchten, in den Sinn kommen, weil sie Expertenwissen in diesem oder einem indirekt relevantem Gebiet wie beispielsweise Informationstechnologie oder klinische Psychologie besitzen, und sind oft sehr nützlich. (Sie zu erhalten ist eines der wertvollen Nebenresultate, die man durch den Einsatz von externen Evaluierenden hat.) Doch Makro-Empfehlungen — welche die Disposition des gesamten Programms betreffen (finanziere, kürze, modifiziere, exportiere etc. — was wir *externe* Managementempfehlungen nennen mögen) — sind normalerweise ein ganz anderes Kaliber. Sie erfordern häufig (i) umfassendes Wissen über den entscheidungsrelevanten Kontext der in Bezug auf das Programm Entscheidungsberechtigten (die nicht immer die Auftraggebenden der Evaluation sind *und* ungewillt oder psychologisch unfähig sein mögen, vollständige Details über den Entscheidungskontext des Programms zu bieten), (ii) nicht unbeträchtlichen Zusatz-Aufwand *und* (iii) Wissen über die Gesamtheit der im Programm verfügbaren Managementoptionen. Hauptbestandteile dieser Liste (i) bis (iii) sind meist nicht jedem verfügbar, Fachpersonen – die grössten Experten unter den Experten eingeschlossen – sind nicht auf dem aktuell neuesten Stand zu diesem bestimmten Thema (das Programm in diesem Kontext zu verbessern) oder sie sind nur für den Vorstand verfügbar oder ausgesuchten Mitgliedern der Legislative oder vielleicht nur deren Psychotherapeuten und, Psychotherapeutinnen daher gänzlich unzugänglich für die Evaluierenden.

Diese Extra-Aufwendungen und manchmal noch eine mehr, gelten auch für die Bereitstellung von Erklärungen bezüglich Erfolg oder Misserfolg. Dieser Mehraufwand liegt zentral darin, über die korrekte (nicht bloss die unterstellte) Programmlogik oder –theorie zu verfügen, was oft mehr — und sehr selten weniger — als die allerneueste Expertise im jeweiligen Feld verlangt. Die so kombinierten Anforderungen bedeuten oft, dass der Versuch, Empfehlungen oder Erklärungen zu liefern, zu Lasten basal gut ausgeführter Evaluationsaufgaben geht (falls man dann überhaupt dazu kommt), was in den meisten Fällen ein fauler Kompromiss ist. Beachten Sie, dass Makro-Empfehlungen typischerweise die Fähigkeit voraussetzen, die Folgen empfohlener Programmveränderungen in diesem bestimmten Kontext vorauszusagen; etwas, das eine Programmtheorie (wie viele sozialwissenschaftliche Theorien) in seltensten Fällen auch nur irgendwie zuverlässig leisten kann. Hingegen sind *prozedurale* Empfehlungen in der Zeitform des Futurs, wie z. B. zu weiterem Bedarf an Forschung oder an Datensammlung oder Evaluation, häufig möglich — dabei weniger nützlich. Einfache *Vorhersagen* sind häufig erwünscht (z. B.: „Wird das Programm mit den empfohlenen Veränderungen zuverlässig funktionieren, ohne dass Veränderungen beim Personal erforderlich sind?“), doch sie sind normalerweise ausgesprochen riskant.

Policy-Analyse, insbesondere bei alternativer Politik die künftig implementiert werden soll, ist im Wesentlichen Programmevaluation von zukünftigen (möglichen) Programmen und schliesst daher notwendigerweise Vorhersagen ein. Umfassendes

25 Die Erfindung des feministischen Slogans „Eine Frau ohne Mann ist wie ein Fisch ohne Fahrrad“ wird vielfach der amerikanischen Journalistin *Gloria Steinem* zugeschrieben; vgl. http://en.wikipedia.org/wiki/Gloria_Steinem [Stand: 02.06.2006].

Wissen und Kenntnis vom Schicksal ähnlicher Programme in der Vergangenheit ist dann eine wesentliche Ressource. Falls die Politik bereits implementiert wurde. Policy-Analyse läuft dann auf etwas sehr Ähnliches wie Programmevaluation hinaus.

Die Tatsache, dass Auftraggebende Erklärungen, Makro-Empfehlungen und Vorhersagen verlangen, ist für uns Anlass, sie darüber aufzuklären, was wir definitiv tun können im Vergleich zu dem, was wir erhoffen, dass es sich als möglich entpuppt. Obwohl verlockend, sind diese Erwartungen keine Entschuldigung dafür, diese Extras durchzuführen, auszuprobieren oder erst recht zu versprechen, wenn Sie nicht über die sehr happigen Extramittel verfügen um dieses zu tun, insbesondere wenn Sie dadurch die primäre Aufgabe der Evaluierenden aufs Spiel setzen, nämlich evaluative Schlussfolgerungen über den Evaluationsgegenstand bereitzustellen. Der kostenfreie bzw. kostenabhängige Wert bzw. die Bedeutung des Programms sind oft schwierig festzustellen. Die Bestimmung, wie der Evaluationsgegenstand verbessert werden kann, warum er funktioniert oder auch nicht, und was man damit machen sollte, sind *andere* Aufgaben, die oft grosses wissenschaftliches und/oder ökonomisches/soziales Interesse wecken, dabei oft über aktuelle wissenschaftliche Kenntnisse geschweige denn über die Fähigkeiten der Evaluierenden, die hoch kompetent sind, ein *Programm* zu evaluieren, hinaus gehen. (M. a. W.: „Black box“-Evaluation sollte nicht als verächtliche Bezeichnung verwendet werden, da sie oft die einfache, sofort erschwingliche und die einzig durchführbare Herangehensweise darstellt.)

13. Fakultative Verantwortung und Rechenschaftslegung

Falls diese festgestellt werden können und falls sie angebracht sind (einige Versionen von Rechenschaftslegung, welche die Rechenschaftslegung von Personen betonen, verlangen dies — siehe untenstehende Beispiele). Die Zuweisung von Tadel und Lob bedarf umfassenden Wissens und Kenntnis über (i) den Kenntnisstand der Hauptakteure zum Zeitpunkt der Entscheidung, (ii) deren Ressourcen und Verantwortlichkeiten sowie (iii) einer ethischen Analyse ihrer Optionen und der Berechtigungen, die sie haben mögen. Nicht viele Evaluatoren und Evaluatorinnen haben die Qualifikationen, um diese Art Analyse durchzuführen. Das „Spiel der Schuldzuweisung“ «*blame game*» ist in den meisten Fällen grundverschieden von Evaluation und man sollte sich nicht leichtfertig darauf einlassen. Dennoch, manchmal werden offenkundige Fehler gemacht, die schwerwiegende Konsequenzen haben, und auf solche sollte hingewiesen werden. Manchmal werden gerechtfertigte Auswahlen mit positiven oder negativen Effekten getroffen und angegriffen und diese sollten im Rahmen der Evaluation gelobt oder verteidigt werden. Die Untersuchungen von Flugzeugunfällen durch das U.S.-amerikanische *National Transportation Safety Board* bilden ein Modellbeispiel wie dies getan werden kann: Dies sind Evaluationen von Ereignissen mit der Erfordernis, Verantwortung zuzuschreiben, egal ob menschliches Versagen oder Naturereignisse. (Todesfälle in Operationssälen stellen ähnliche Probleme dar und werden manchmal genauso intensiv untersucht.)

ANMERKUNG: Die Evaluation von Katastrophen, in jüngster Zeit ein Gebiet mit beachtlicher Aktivität, umfasst normalerweise einige oder alle der folgenden fünf Komponenten: (i) Evaluation des Ausmasses der Bereitschaft für den Katastrophenfall; (ii) Evaluation der unmittelbaren Reaktion; (iii) Evaluation der Hilfsaktionen; (iv) Evaluation der gelernten Lektionen; (v) Evaluation von Korrekturmassnahmen. Alle fünf leisten ein Stück Evaluation der Verantwortlichkeit sowie Zuweisung von Lob und Tadel. Jüngste Bestrebungen scheinen diese Komponenten nicht sauber zu unterscheiden oder alle von ihnen abgedeckt zu haben.

14. Bericht und Unterstützung

Jetzt kommen wir zur Aufgabe, die Schlussfolgerungen in angemessener Weise und zu angebrachter Zeit zu vermitteln. Das ist eine ganz andere Aufgabe als — doch sie wird häufig verwechselt mit — der Übergabe eines für Forschungsstudien typischen Schlussberichts, der kommentierende Darstellungen mit Tabellen, Grafiken und anderen technischen Elementen enthält. Berichten in der Evaluation kann an verschiedene Adressaten und Adressatinnen in verschiedenen Evaluationsphasen radikal unterschiedliche Anforderungen an Präsentationen stellen: Diese können mündlich oder schriftlich, lang oder kurz, öffentlich oder privat, technisch oder nicht technisch, grafisch oder verbal, wissenschaftlich oder erzählerisch, anekdotisch oder stichpunktartig sein. Hilfen zum Bericht sollten eingeschlossen sein, z. B. sollten später oder sofort aufkommende Fragen beantwortet werden. Ausserdem sollte die Relevanz des Berichtes für verschiedene Gruppen, einschliesslich Nutzer und Nutzerinnen, Personal, Geldgebende und andere *impactees*, erläutert werden. Dies kann die Erstellung und Beschreibung von verschiedenen möglichen Szenarien umfassen, die mit den Ergebnissen übereinstimmen oder von ihnen abweichen, d. h. die Auftraggebenden werden bei der Problemlösung unterstützt. Diverse Kommunikationsfähigkeiten sind für diese Prozesse häufig nützlich, z. B. Reaktionen des Publikums verstehen und darauf reagieren können, Einsatz und Entschlüsseln von Körpersprache, Verstehen kultureller Bilddeutung «*iconography*» und der konnotativen Implikationen von Präsentationen und Reaktionen (die „konnotativen Implikationen“ machen den subexpliziten aber supra-symbolischen Kommunikationsbereich aus, beispielsweise geschlechtssensible oder geschlechtsunsensible Sprache). Unternehmen Sie einen expliziten Versuch, gelernte Lektionen «*lessons learned*» zu identifizieren. Checkpunkt 14 sollte ausserdem abdecken, die Ergebnisse (und auch beiläufige Erkenntnisse), in relevante Datenbanken einzustellen, falls diese vorhanden sind. Möglichst, aber nicht unbedingt, sollte in Fachzeitschriften publiziert werden. Der Aufbau einer neuen Datenbank oder eines Informationskanals (z. B. eines Newsletters) sollte empfohlen werden, soweit nutzenversprechend. Schliesslich sollte die Veröffentlichung über weitere Kanäle erwogen werden, z. B. mittels Präsentationen, durch das Internet, Diskussionen auf wissenschaftlichen Tagungen oder als Hardcopy Poster, Plakatanschlüge, Fahrzeug- und Grossflächenwerbung sowie Filme.

15. Meta-Evaluation

Dies ist die Evaluation einer Evaluation, was die mit den vorangehenden Checkpunkten bezeichneten Schritte einschliesst, um Stärken/Beschränkungen/andere Verwendungsmöglichkeiten zu identifizieren. Die Meta-Evaluation sollte immer als separater Schritt der Qualitätskontrolle ausgeführt werden, nach folgenden Punkten:²⁶ (i) soweit wie möglich, durch den Evaluierenden/die Evaluierende, gewiss — doch nicht ausschliesslich — nach der Fertigstellung des Abschlussberichtes und (ii) sooft wie möglich *auch* durch eine(n) gegenüber dieser Evaluation extern Evaluierende/-n (der/die Meta-Evaluierende). Hauptkriterien für die Qualität/Güte von Evaluationen sind: (i) Validität zusammen mit (ii) Nützlichkeit (normalerweise für Auftraggebende, Adressaten und Adressatinnen sowie Stakeholder); (iii) Glaubwürdigkeit (für ausgewählte Stakeholder, insbesondere Geldgebende, Aufsichtsbehörden und normalerweise auch Programmpersonal); (iv) Kostenwirksamkeit und (v) ethische Kriterien. Es gibt fünf mögliche Herangehensweisen. Sie und die Meta-Evaluierenden können (a) die KEC auf die Evaluation anwenden und/oder (b) spezielle Meta-Evaluations-Checklisten verwenden (mehrere sind verfügbar) und/oder (c) die Evaluation, so wie sie durchgeführt wurde, replizieren und die Ergebnisse vergleichen und/oder (d) die

26 Die KEC kann zur Selbstkontrolle dienen, vorzugsweise ist sie jedoch durch extern Metaevaluierende einzusetzen.

Evaluation bei Verwendung einer anderen Methodologie durchführen und Ergebnisse vergleichen und/oder (e) die Standards für Programmevaluationen verwenden. Es ist sehr wünschenswert, mehr als eine der Herangehensweisen anzuwenden.

ANMERKUNGEN: (a) Nützlichkeit ist Nutzbarkeit und nicht die tatsächliche Nutzung; letztere — oder deren Fehlen — ist bestenfalls ein schwacher Indikator des vorhergehenden. (b) Implementierung beweist nicht hohe Nutzbarkeit. (c) Wortgetreue oder direkte Nutzung sind auf Evaluationen ohne Empfehlungen nicht ohne weiteres anwendbar, eine Kategorie, die viele wichtige, vollständige und einflussreiche Evaluationen einschliesst: „Verbindlich in Erwägung ziehen“ «*due consideration*» ist eine bessere Bezeichnung für die ideale Reaktion. (d) Die Auswirkungen von Evaluationen treten häufig erst Jahre nach Abgabe des Berichtes auf oder sogar wenn die Evaluation völlig abgelehnt wurde. (e) Unterstützung zur Nutzbarmachung der Evaluation über den Abgabezeitpunkt des Berichtes hinaus sollte zumindest angeboten werden. (f) Identifizieren Sie Beiträge der Evaluation zum Wissensmanagementsystem der Organisation; falls kein System vorhanden ist, empfehlen Sie dessen Aufbau. (g) Bedenken Sie, dass die Effekte der Evaluation keine Effekte des Programms sind. Eine selbstorganisations-gesteuerte Evaluation («*empowerment evaluation*») löst beim Programmpersonal einen substanziellen Wissenszuwachs über Evaluation aus und erhöht ihre Bereitschaft, Evaluationen zu nutzen. Jedoch ist dies aus Sicht der Evaluierenden kein relevanter Effekt des Programms. Obwohl dieser wertvolle Outcome ein Effekt der Evaluation ist, kann er geringe Validität/externe Glaubwürdigkeit nicht kompensieren, da er kein Hauptkriterium für die Qualität/Güte des Evaluationsgegenstandes ist. Dies sind weit verbreitete Probleme selbstorganisations-gesteuerter Evaluationen. (h) Gleichermassen sind die üblichen nicht-geldlichen Kosten der Evaluation — Arbeitsunterbrechungen für das Personal — *keine* negativen Effekte des Programms; und natürlich treten sie bei zielfreien Evaluationen nur minimal ein, da die Evaluierenden nicht mit dem Personal sprechen (*einer* der Punkte, der in der Meta-Evaluation aufgegriffen werden sollte). Ein sorgfältiges Evaluationsdesign kann solche Kosten manchmal fast auf Null reduzieren oder Vorteile bieten, welche die Kosten mehr als ausgleichen.

HAUPTANMERKUNG D: Die beiden als „möglich“ markierten Checkpunkte 12 und 13 sind nicht immer relevant oder praktikabel durchführbar und verlangen häufig zusätzlich Zeit/Kosten. Sie sind aufgeführt, weil sie oft als obligatorische oder „unhinterfragte“ Bestandteile für professionelle Evaluation unterstellt werden, was nicht zutrifft: „Black box–Evaluation“ ist oft nützlich und die einzige realisierbare Variante und/oder sie ist die beste Herangehensweise, um die Fragen der Auftraggebenden zu beantworten (gewiss ist es die schnellste und kostengünstigste Variante, was ihr oft einen Vorsprung gegenüber der besten Variante verschafft). Die Ausführungen zu den beiden Checkpunkten bieten Orientierungshilfen, falls praktikabel durchführbar und von den Auftraggebenden erwünscht.

HAUPTANMERKUNG E: Diese erklärenden Bemerkungen sollen als Annäherungen zum Inhalt jedes Checkpunkts dienen und nichts mehr. Weitere Details zu den Anmerkungen und den einzelnen Punkten sind in *Michael Scrivens* (4th edition, 1991) *Evaluation Thesaurus* (unter „Key Evaluation Checklist“ oder dem spezifischen Eintrag) oder in zitierten Referenzen oder in *E. Jane Davidsons* (2005) *Evaluation Methodology Basics* als gegenwärtig bester Quelle nachzulesen. Die vorliegende Version der KEC ist dank vieler Studierender und Kollegen/Kolleginnen, aus jüngster Zeit - genannt seien insbesondere *Emil Posavac*, *Jane Davidson*, *Rob Brinkerhoff*, *Lori Wingate* und *Andrea Wulf* sowie ein oder zwei Gedanken aus *Michael Quinn Patton's* Arbeit - gegenüber der im *Evaluation Thesaurus* deutlich verbessert. Weitere Anregungen und Kritik sind willkommen — bitte senden Sie diese an:

michael.scriven@wmich.edu



Literatur

- Beywl, Wolfgang (1988). *Zur Weiterentwicklung der Evaluationsmethodologie*. Frankfurt am Main: Lang.
- Deutsche Gesellschaft für Evaluation (2002). *Standards für Evaluation*. Köln: DeGEval.
- Davidson, Jane E. (2005). *Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation*. Thousand Oaks, CA: Sage.
- Joint Committee on Standards for Educational Evaluation (Chair: James Sanders) (1994). *The Program Evaluation Standards*. 2nd ed. Thousand Oaks, CA: Sage.
- Sanders, James (Hrsg.) (2006). *Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation, 3., erw. und akt. Aufl., Wiesbaden: VS Verlag*.
- Scriven, Michael (2004). *Editorial: The Fiefdom Problem*. Journal of MultiDisciplinary Evaluation 1. Frei verfügbar unter: <http://evaluation.wmich.edu/jmde/>
- Scriven, Michael (1991). *Evaluation Thesaurus*. 4th edition. Newbury Park, CA: Sage.
- Scriven, Michael (1998). „Minimalist Theory: The Least Theory That Practice Requires.“ American Journal of Evaluation, 19(1). 57-70.
- Scriven, Michael (2005). *The Logic and Methodology of Checklists*. Frei verfügbar unter: http://www.wmich.edu/evalctr/checklists/papers/logic&methodology_oct05.pdf
- Scriven, Michael with Jane Roth (1977). *Evaluation Thesaurus*, 1st. edition. Edgepress [darin: Erstveröffentlichung der KEC].
- Widmer, T., Landert, C., & Bachmann, N. (2000). *Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards)*. Genève/Bern: SEVAL.

Disclaimer

Die Checkliste ist Bestandteil der *Evaluation Checklists Website* des *Evaluation Center* an der *Western Michigan University*. Die Nutzenden erhalten sie als kostenlosen Service. Die Verfassenden haben die Checklisten nicht auf einen spezifischen Bedarf zugeschnitten; die Nutzenden wenden die Checklisten nach eigenem Ermessen und Urteil an. Es werden keine Zusicherungen oder Garantien gegeben, dass diese Checkliste für den spezifischen Zweck geeignet ist, für den die Nutzenden sie einsetzen wollen. Die Verfassenden stellen solche Garantien oder Zusicherungen ausdrücklich in Abrede.

Diese deutschsprachige Fassung erscheint auf der Website von Univation.

Die Übersetzungen erfolgen mit Erlaubnis der Verfassenden; mehrdeutige Textstellen sind teils in Rücksprache mit ihnen geklärt. Teils finden sich Erläuterungen der Übersetzenden in Fussnoten. Bei der Übersetzung von Akteurs-Begriffen wie *participant* werden geschlechtsneutrale Bezeichnungen bevorzugt (z. B. „Teilnehmende“) oder beide grammatikalischen Geschlechter genannt (z. B. „Kunden und Kundinnen“). Gelegentlich wird im Plural formuliert, auch wenn im Original der Singular steht. In Ausnahmefällen wird zur Vereinfachung des Satzaufbaus/bei Platzmangel (in Fragebogen, Tabellen ...) das generische Maskulinum genutzt (z. B. „Einwohnerzahlen“). Teils werden den übersetzten Begriffen die englischsprachigen Originalbegriffe in doppelten spitzen Klammern und kursiv ausgezeichnet nachgestellt. Beispiel: Auftraggebende «clients».

Mehr Informationen zur Zwecksetzung und Konzeption der Checklisten finden sich auf:

www.wmich.edu/evalctr/checklists/

Besonders hingewiesen sei auf die Leitlinien für eine akzeptable Verwendung der Checklisten:

www.wmich.edu/evalctr/checklists/policy.html

Copyright beim Verfasser; für die Übersetzung ins Deutsche zusätzlich bei den Übersetzenden. **Zitierweise:** Scriven, Michael (1999): *Die Key Evaluation Checklist*. 2006 übersetzt und kommentiert von Daniela Schröter und Wolfgang Beywl. Downloadbar von www.univation.org/checklisten

