

# Was kennzeichnet gute Evaluationen in der Berufsbildung?

## Standards für Evaluation und Perspektiven ihrer Weiterentwicklung

► Die Deutsche Gesellschaft für Evaluation hat allgemeine Evaluations-Standards formuliert, die im Beitrag erläutert und deren Anwendbarkeit auf die Berufsbildung diskutiert wird. Es werden Thesen für die Weiterentwicklung der Qualität von Evaluationen in der beruflichen Bildung vorgestellt. Für die systematische Entwicklung und Überprüfung von Qualität in der beruflichen Bildung werden seit geraumer Zeit „Qualitätsmanagement“, „Qualitätssicherung“ oder auch Verfahren, die dem vergleichenden Warentest nahe stehen, intensiv diskutiert. Hingegen steht die Auseinandersetzung über „Evaluation“, die bereits seit Jahrzehnten in der Berufsbildung als sozialwissenschaftlich begründete Vorgehensweise genutzt wird, in Deutschland wie in den anderen Mitgliedsländern der EU am Anfang.

Ein umfangreicher Fundus an Forschungsberichten, Lehrbüchern, web-basierten Quellen und Diskussionsforen bis hin zu Studiengängen zum Thema Evaluation im Bildungsbereich findet sich in Nordamerika. So hat das US-amerikanische „Joint Committee on Standards for Educational Evaluation“ (JC) 1994 bereits zum zweiten Mal Qualitätsleitlinien für die Evaluation im Bereich von Bildung und Erziehung veröffentlicht.<sup>1</sup> Da die Standards im amerikanischen Ursprung dem Bildungsbereich entstammen, liegt es nahe, dass sie grundsätzlich auch für den Bereich der beruflichen Bildung Gültigkeit haben. Anhaltspunkt dafür ist auch, dass sich im Selbstlernmaterial zu den JC-Standards allein sieben Anschauungsbeispiele aus dem Umfeld der betrieblichen Bildung finden.

Die 1997 gegründete Deutsche Gesellschaft für Evaluation (DeGEval) hat auf die Arbeiten des Joint Committee aufbauend „Standards für Evaluation“ formuliert.<sup>2</sup> Professionelle Evaluationen sollen demnach vier grundlegende Eigenschaften aufweisen:

- Nützlichkeit,
- Durchführbarkeit,
- Fairness und
- Genauigkeit.

Diesen vier Standardgruppen sind insgesamt 25 Einzelstandards zugeordnet. Die DeGEval-Standards erfüllen als „Maximal-Standards“ eine Orientierungsfunktion und wollen zum Dialog über die Qualität von Evaluationen anregen.

Im Rahmen eines auf die europäischen Mitgliedsstaaten ausgerichteten Forschungsvorhabens<sup>3</sup> wurde folgende Fragestellung untersucht: Können die auf breite Anwendungsfelder gerichteten, auf die nordamerikanische Tradition aufbauenden DeGEval-Standards auch in institutionell und curricular unterschiedlich geprägten Berufsbildungssystemen Geltung beanspruchen?

Im Ergebnis sind nicht alle Einzel-Standards in jedem Evaluationsprojekt gleichermaßen anwendbar. Durch die Lite-



**WOLFGANG BEYWL**

Dr. päd., M. A., Univision, Institut für Evaluation, Köln, wissenschaftlicher Leiter und Leiter der postgradualen Studiengänge „Evaluation“, Universität Bern



**SANDRA SPEER**

Dipl.-Volkswirtin, Univision, Institut für Evaluation, Köln, Associate

raturanalyse wurde jedoch die Anwendbarkeit jedes einzelnen DeGEval-Standards in unterschiedlichen Kontexten der Berufsbildung bestätigt. Es finden sich in der untersuchten Literatur konkrete Qualitätsansprüche, Ratschläge und Anwendungshinweise für Evaluationen, wie sie in den einzelnen DeGEval-Standards inbegriffen sind. Die einzelnen Standards können auf diese Weise sinnvoll im Bezug auf den Evaluationsgegenstand der Berufsbildung und seine Besonderheiten hin illustriert werden, was im Folgenden aufgezeigt wird.

**Nützlichkeit**

„Die Nützlichkeitsstandards sollen sicherstellen, dass sich die Evaluation an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet“ (DeGEval 2002, S. 8).

Nützlichkeit spielt für die (intendierten) Nutzer/-innen von Evaluationen und ihren Ergebnissen eine herausragende Rolle. In der Berufsbildung zählen zu den „Beteiligten und Betroffenen“ (N1)<sup>4</sup> neben den Trägern und Verantwortlichen für die Programme und Maßnahmen insbesondere Arbeitnehmer- und Arbeitgebervertretungen. In der Vorbereitung einer Evaluation müssen die Verwendungszwecke der Evaluation geklärt (N2) werden: Entweder geht es um die Vorbereitung von Entscheidungen zur Berufsbildung, sei es von Regierung, Bildungseinrichtungen oder Unternehmen, oder es sollen empirische Grundlagen für die Optimierung von Maßnahmen der Berufsbildung geschaffen werden.

Evaluatorisches Know-how ist bereits bei der Planung von Berufsbildungsprogrammen relevant: Hier geht es um die Klärung und Formulierung von Programmzielen, deren Erreichen später überprüft, gemessen werden soll. Frühzeitig sollten empirische Daten zur Ausgangssituation der Bildungsteilnehmenden erhoben, Vorbereitungen für Vergleichs- oder Kontrollgruppen getroffen werden.

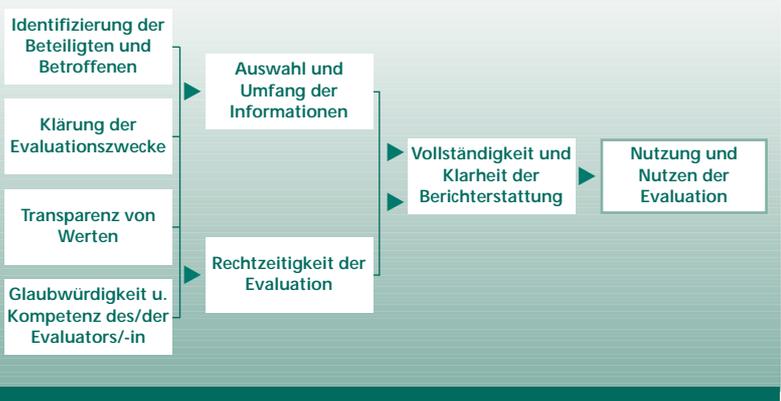


Abbildung 1 Sieben Sachen, die Evaluationen nützlich machen

Evaluationsberichte sollten zielgruppenspezifisch erstellt werden, so dass wichtige Erkenntnisse und Ergebnisse für die Berufsbildungspolitik, -forschung und -praxis nutzbar gemacht werden können. Zwischenpräsentationen, Expertenworkshops oder Pressematerial können dafür genau so wichtig sein wie ein abschließender Bericht, insbesondere im Rahmen von formativen Evaluationen. Die Gruppe der möglichen Nutzer/-innen einer Evaluation ist möglichst weit zu fassen, indem z. B. nicht nur Verantwortliche und Teilnehmende einer Trainingsmaßnahme, sondern auch weitere für die Berufsbildung verantwortliche Mitarbeiter/-innen eines Unternehmens informiert werden.

**Durchführbarkeit**

„Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird“ (DeGEval, 2002, S. 9).

Sowohl die Angemessenheit der eingesetzten Erhebungsverfahren (D1) als auch das diplomatische Vorgehen (D2) beeinflussen maßgeblich die Effizienz von Evaluationen (D3).

Evaluationen können einerseits Programme als Gegenstand haben, die relativ klein, auf eine bestimmte Zielgruppe fokussiert und zeitlich befristet sind, z. B. eine betriebliche Weiterbildungsmaßnahme für einen begrenzten Teilnehmerkreis. Im anderen Fall sind es von der Bundesebene gemeinsam mit den Ländern koordinierte, durch Trägerverbände umgesetzte mehrjährige Programme, die an hunderterten von Standorten stattfinden und die gleichzeitig mehrere kaskadenartig verkoppelte Zielgruppen haben (z. B. Geschäftsführer/-innen von Trägern, Ausbilder von Ausbildern, Ausbilder, End-Zielgruppe Jugendliche mit ihren Eltern). Für verschiedenste Evaluationsgegenstände angemessene Untersuchungsdesigns und -methoden zu bestimmen, macht die Evaluationsplanung zu einer komplexen, langjährige Erfahrungen fordernden Aufgabe. So können bei „großen“ Programmen quantitative, standardisierte Instrumente, für die ein erheblicher Entwicklungs- oder Anpassungsaufwand geleistet werden muss, effizient sein. Hingegen liegen für kleine, noch in Entwicklung befindliche Programme qualitative, in der konkreten Anwendung flexibel anpassbare Instrumente nahe. In der Regel müssen qualitative und quantitative Methoden kombiniert werden.

Die Praxis zeigt, dass Evaluationsprojekte am Widerstand von Leitungspersonen oder der Arbeitnehmervertretung scheitern können. Einzubeziehen ist auch der Datenschutz. Evaluation trifft auch dann auf Ablehnung, wenn der Untersuchungsapparat zu groß dimensioniert wird oder die Lehrenden und Lernenden mit Erhebungen überlastet wer-

den. Bei Verbundprojekten oder Programmen mit erheblicher politischer Bedeutung ist Verhandlungsgeschick und eine klares Informationswesen der Evaluation entscheidend für ihre Akzeptanz bei den Entscheidungsträgern, welche als Nutzer der Evaluationsergebnisse erreicht werden sollen. Bei der Budgetierung einer Evaluation sollte berücksichtigt werden, welche Tragweite für ihre Ergebnisse angestrebt ist. So kann z. B. für die Evaluation eines Pilotprojekts eine in Relation zu den Programmkosten hoher Kostenanteil sinnvoll sein, da bei einer späteren Übertragung auf die Breite der Berufsbildung Fehler vermieden und so erhebliche Kosteneinsparungen erzielt werden können. In der Praxis wird es hier oftmals zu Kompromissen zwischen dem Anspruch an methodische Perfektion (siehe auch Genauigkeits-Standards) und verfügbaren finanziellen und zeitlichen Ressourcen kommen.

### Fairness

„Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird“ (DeGEval, 2002, S. 9).

Die ersten beiden Fairnessstandards „Formale Vereinbarung“ (F1) und „Schutz individueller Rechte“ (F2) haben klare Bezugspunkte z. B. in Form von Musterverträgen oder gesetzlichen Grundlagen. Die vollständige und faire Überprüfung (F3) sowie die unparteiische Durchführung und Berichterstattung (F4) hingegen stehen in einem permanenten Spannungsfeld verschiedenster Interessen. Auch die „Offenlegung der Ergebnisse“ (F5) ist hiervon berührt: Werden alle relevanten Ergebnisse offen gelegt oder ausschließlich solche, die nicht substantiell mit den Interessen zentraler Beteiligter kollidieren? Solche „politischen“ Entscheidungen sollten frühzeitig im Ablauf einer Evaluation getroffen, formal vereinbart (F1) und gegenüber den Beteiligten und Betroffenen kommuniziert werden (vgl. D2).

In den Fairnessstandards werden die Anforderungen an Evaluationen in der Berufsbildung aus „Industrial Relations“-Sicht, wie Mitbestimmungsregelungen und Datenschutz, besonders deutlich. Die Fairnessstandards stellen außerdem besondere Anforderungen an die rechtlichen Kenntnisse und die soziale Sensibilität des Evaluators (vgl. N3). Dem Schutz personenbezogener Daten und dem angemessenen Umgang mit Leistungs- und Ergebnisdaten, die auf einzelne Personen rückführbar sein könnten, gebührt daher besonders in Berufsbildungs-Evaluationen höchste Aufmerksamkeit (z. B., wenn Intelligenz oder andere Persönlichkeitsfaktoren von Maßnahmenteilnehmenden bestimmt werden).

Für die berufliche Bildung ist kennzeichnend, dass an ihr zahlreiche Interessengruppen mitwirken. Evaluationen

Abbildung 2 Zwei Wege, die Evaluationen durchführbar machen

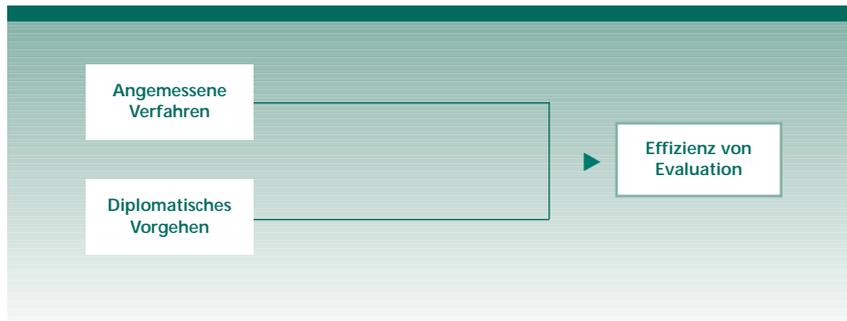


Abbildung 3 Fünf Leitplanken, die Evaluationen auf fairem Kurs halten

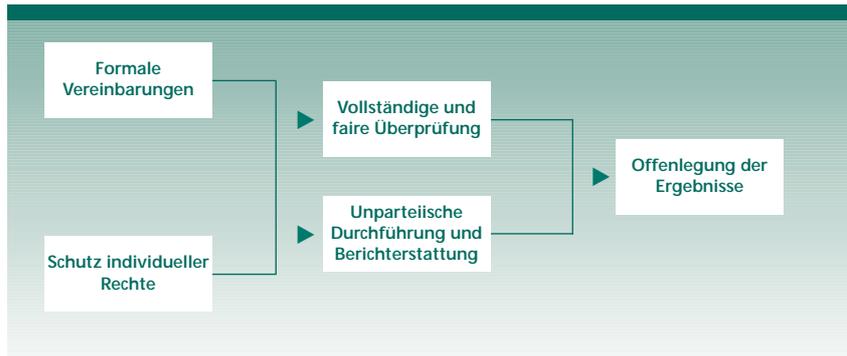
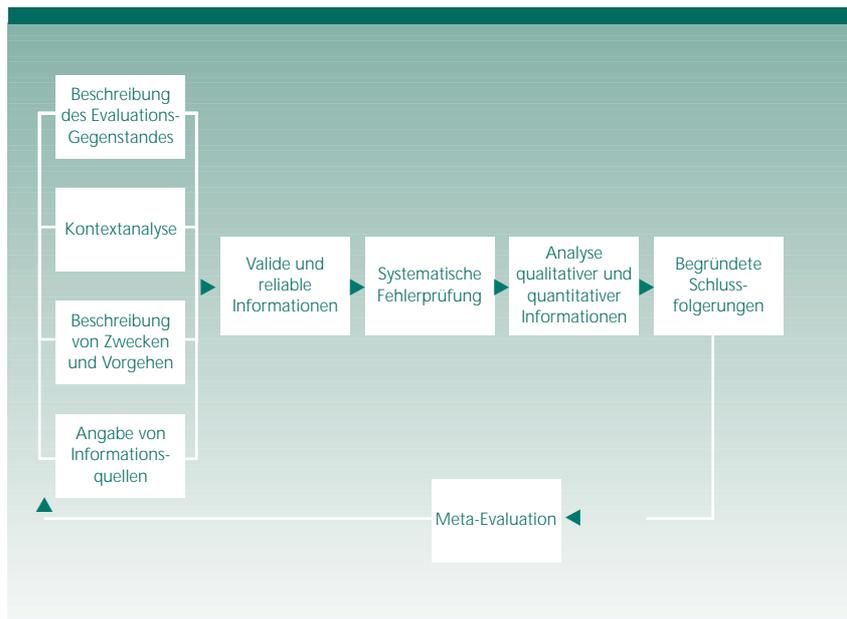


Abbildung 4 Neun Bausteine, die Evaluationen genau machen



müssen deren Perspektiven angemessen einbeziehen, was in der Untersuchungspraxis bedeutet, sie fortlaufend zu informieren und zu beteiligen. Evaluation in einem durch multiple Stakeholder geprägten Feld wie z. B. den Ausbildungspartnerschaften oder den Modellversuchen in der schulischen und außerschulischen Bildung kann dann nutzbringend sein, wenn sie die Informationsbedürfnisse der verschiedenen Beteiligten aufnimmt und ihre Annahmen über den zu bewertenden Bildungsbereich einer nachvollziehbaren empirischen Überprüfung unterwirft. Unab-

hängigkeit von Evaluation bedeutet nicht, die Werte und Interessen der Beteiligten zu ignorieren, sie durch wissenschaftliche Theorien oder technische Regelwerke zu ersetzen, sondern die Übereinstimmungen und Konflikte sichtbar zu machen, die Möglichkeiten empirischer Nachweise auszureizen und gleichzeitig deren Grenzen – bei unvereinbaren Interessenpositionen – klar zu benennen.

Schließlich ist die geforderte Unparteilichkeit dann besonders kritisch zu prüfen, wenn – wie in Modellprojekten üblich – der Evaluator/die Evaluatorin das Programm mit entwickelt, seine Umsetzung formativ begleitet und schließlich auch seine Resultate und Wirkungen beschreibt und bewertet. Dies führt zu Rollenkonflikten, deren Beherrschung höchste Anforderungen an die Professionalität stellt.<sup>5</sup>

## Genauigkeit

„Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt“ (DeGEval, 2002, S. 10).

Die neun Standards dieser Gruppe lassen sich wie folgt gliedern: Die ersten zwei Standards (G1 und G2) sprechen die Bestimmung des Evaluationsgegenstandes in seinem Kontext an. G3 und G4 fordern die genaue Dokumentation von Zwecksetzung, Vorgehen und der in der Evaluation genutzten Informationsquellen. Die folgenden vier Standards beziehen sich auf den eigentlichen Prozess der Erhebung, Prüfung, Auswertung und schließlich Nutzung der Daten für die Formulierung von Schlussfolgerungen – hier sind Anforderungen an die Gewinnung von Informationen und Befunden formuliert. Mit dem Standard G9 wird Metaevaluation als Methode der Qualitätssicherung und Qualitätsentwicklung für Evaluationen selbst gefordert.

Einige Aspekte der Standardgruppe „Genauigkeit“ werden in der Literatur zu Qualitätsanforderungen an Evaluationen in der Berufsbildung kaum diskutiert. Die Standards zur empirischen Informationsgewinnung entsprechen den Kriterien, die standardmäßig für die Qualität von wissenschaftlichen Untersuchungen gefordert werden. Diese Kriterien werden in sozialwissenschaftlichen Lehrbüchern ausführlich behandelt und vielleicht deshalb als so selbstverständlich angesehen, dass sie nicht eigens angesprochen werden. Allerdings ist zu bedenken, dass Evaluationen nicht ausschließlich von Spezialisten/-innen für empirische Methoden durchgeführt werden und dass auch Auftraggeber von Evaluationen mit den Standards in die Lage versetzt werden sollen, sich ein Bild von „guter Evaluation“ zu machen.<sup>6</sup> Da empirisch-wissenschaftliche Methoden im Rahmen von Evaluationen in Abwägung mit den Anforderungen nach Nützlichkeit und Durchführbarkeit eingesetzt werden, müssen deren aus der Grundlagenforschung stammenden Qualitätsmaßstäbe angepasst werden.

In der Literatur zur Berufsbildung fällt auf, dass die Genauigkeitsstandards mehr für mikro-/und makroökonomische Fragestellungen als für Evaluationen im Bildungsträger- und Unternehmensbereich thematisiert werden. Im ersten Fall steht vielfach die Generalisierbarkeit der Ergebnisse im Vordergrund, während sich die Anwendung im letzten Fall oft auf den Träger der Bildungsmaßnahme beschränkt.

## Perspektiven

- Die Evaluations-Standards können die Qualität von Evaluationen verbessern und ermöglichen es, das Verfahren transparent zu machen und nach außen zu legitimieren.
- Für Evaluationen in der Berufsbildung gibt es in den Standardgruppen Nützlichkeit, Durchführbarkeit und Fairness deutlich mehr Anknüpfungspunkte für Spezifi-

### Anmerkungen

- 1 Vgl. als deutschsprachige Übersetzung: Joint Committee on Standards for Educational Evaluation (JC) (Hrsg.): *Handbuch der Evaluationsstandards*, Opladen 2000
- 2 Deutsche Gesellschaft für Evaluation e. V. (DeGEval): *Standards für Evaluation*, Köln 2002. (siehe auch: [www.degeval.de/standards/Standards.pdf](http://www.degeval.de/standards/Standards.pdf))
- 3 Vgl. ausführlich in Beywl, W.; Speer, S.: *Standards for Evaluation Practices. On the Way to Developing Standards for Programme Evaluation in Vocational Education and*

*Training Contexts*. In: Descy, P.; Tessaring, M. (Hrsg.): *Third Report on Vocational Training Research in Europe: Background Report*, CEDEFOP, Luxembourg 2003 [im Erscheinen]. Im Rahmen der Literaturrecherche wurden einschlägige Beiträge der letzten fünf Jahre ausgewertet, die eine explizite Reflexion der Qualität von Evaluationen bzw. Anforderungen an Evaluationen im Bereich der Berufsbildung leisten.

- 4 Die Standards werden durch einen Buchstaben wie bspw. „N“ für die Standardgruppe

„Nützlichkeit“ sowie eine folgende Nummer für die Kennzeichnung des Einzel-Standards abgekürzt.

- 5 Sicher wäre dieser Konflikt zu vermeiden, wenn das Evaluationsteam im Übergang von der formativen zur summarischen Phase ausgetauscht würde, was jedoch die Kosten der Evaluation erhöhen würde.
- 6 Es gibt auch eigene Standard-Sets zu diesen Genauigkeits-Anforderungen, die von nationalen berufsständischen Gruppen, wissenschaftlichen Vereinigungen oder auch von

Forschungsförder-Einrichtungen aufgestellt worden sind. Vgl. Beywl, W.; Widmer, Th.: *Die „Standards“ im Vergleich mit weiteren Regelwerken zur Qualität fachlicher Leistungserstellung*. In: Joint Committee on Standards for Educational Evaluation (JC) (Hrsg.), *Handbuch der Evaluationsstandards*, Opladen 2000, S. 259–295

- 7 Vgl. Beywl W.; Müller-Kohlenberg, H.: *Standards der Selbstevaluation, Begründung und aktueller Diskussionsstand*. In: *Zeitschrift für Evaluation*, 1/2003, S. 65–75

zierungen als bei den Genauigkeitsstandards, die vielfach allgemein gültige Anforderungen an empirische Untersuchungen formulieren.

- Als dauerhaftes Spannungsfeld bestätigt sich auch für Berufsbildungsevaluationen die Polarität zwischen „Genauigkeit“ und „Nützlichkeit“. Bisweilen stehen sich die Erwartungen nach unmittelbar nutzbaren Ergebnissen und die Anforderungen aus der empirischen Sozial- und Wirtschaftsforschung – z. B. bezüglich Validität und Reliabilität – kaum vereinbar gegenüber.
- Da besonders das Instrument der Selbstevaluation in Berufsbildungseinrichtungen stark nachgefragt ist, sollte für DeGEval-Standards herausgestellt werden, dass diese vorrangig für interne und externe, arbeitsteilig von Evaluationsspezialisten durchgeführte Untersuchungen gelten. Für Selbstevaluationen sollte ein angepasstes Standardset genutzt werden.<sup>7</sup>
- Ein zusätzlicher berufsbildungsspezifischer Standard „Unterstützung der Qualitätsorientierung in der Beruflichen Bildung“ könnte lauten: „Evaluationen sollen die für Berufsbildungspolitik und -programme Verantwortlichen dabei unterstützen, Qualitätsanforderungen an die berufliche Bildung einzuhalten.“ Im Kommentar wäre anzusprechen: eine Abstimmung auf die Bedarfe der Zielgruppen, der Sozialpartner und der Gesellschaft, die Forderung nach einer wissenschaftlich begründeten inhaltlichen und didaktischen Konzeption, die Ausgestal-

tung von Strukturen und Organisationen der beruflichen Bildung, die Steuerung der Bildungsprozesse und der Wirtschaftlichkeit von Berufsbildungs-Leistungen. Es sollten verbreitete und anerkannte Berufsbildungs-Standards genannt werden.

- Ein zusätzlicher allgemeiner Evaluationsstandard sollte die Evaluatoren und Evaluatorinnen auffordern, das Evaluationsmodell zu explizieren, das sie für ihre konkrete Evaluation nutzen, und die Entscheidung zu begründen. Eine solche Offenlegungs- und Begründungsanforderung könnte die Verbreitung von Evaluationsmodellen und die Diskussion ihrer Schwächen und Stärken sowie eine Kultur der Metaevaluation in Deutschland unterstützen.
- Für die Evaluation gibt es in Deutschland trotz steigender Nachfrage kein Berufsbild. In der universitären Ausbildung wird Evaluation kaum angeboten. Ein systematischer Aufbau von Qualifikationen ist wünschenswert.
- Es wird empfohlen, speziell für das Evaluationsfeld der Berufsbildung begleitende Materialien zu den Evaluationsstandards zu entwickeln, die insbesondere die kritische Analyse von Anschauungsbeispielen durchgeführter Evaluationen auf möglichst vielen relevanten Systemebenen enthält wie z. B. selbst gesteuertes Lernen, Unternehmen/Schule, Lernortverbünde, Kommunen und Regionen, nationale und europaweite Berufsbildungs-Programme. ■