

Evaluationsstandards in der (internationalen) Berufsbildungsforschung

SANDRA BOHLINGER, HOANG LONG NGUYEN (TU DRESDEN)

Abstract

Die Nützlichkeit und Akzeptanz von Evaluationen stehen und fallen mit ihrer Qualität, für deren Sicherung wiederum Standards benötigt werden. Angesichts dessen, dass in Evaluationsvorhaben der beruflichen Bildung eine begrenzte Umsetzung der Evaluierungen auf Basis konkreter Standards moniert wird, ist es Ziel des Beitrags, (internationale) Evaluationsstandards vorzustellen, miteinander zu vergleichen und hinsichtlich ihres Einsatzes in der Berufsbildungsforschung und -zusammenarbeit zu untersuchen. Hierbei handelt es sich um die Standards der DeGEval, der OECD-DAC sowie der UNEG.

1 Einleitung

Evaluationen sollen helfen, Unsicherheiten und Unklarheiten zu überwinden. In einer Zeit, die von „Ungewissheitsgewissheit“ (Marchart 2013, S. 26) sowie technologischen und sozialen Transformationen geprägt ist, sollen sie möglichst zuverlässige Daten liefern und somit die Zukunft ein Stück weit planbarer machen. Evaluation als Teil der Qualitätssicherung (Gonon 2008, S. 96) bietet im Kontext der Berufsbildungsforschung zugleich einen Ansatz, um möglichst stabile Rahmenbedingungen für die Entwicklung von Qualifikationen und Kompetenzen durch unterschiedliche Normen und Standards zu schaffen.

Evaluation ist keineswegs ein neues Thema der Berufsbildungsforschung. Bereits vor rund 20 Jahren attestierte der damalige stellvertretende Generalsekretär und Leiter des Forschungsbereichs des deutschen Bundesinstituts für Berufsbildung, Walter Brosi, der Evaluationsforschung eine zunehmende Bedeutung, „sei es im Rahmen der Modellversuche, der Evaluation berufsbildungspolitischer Programme und Maßnahmen, der Evaluation neuer Prüfungsformen oder der Evaluation modernisierter Ausbildungsberufe“ (Brosi 2003, S. 3). Richtete sich das Hauptaugenmerk damaliger Evaluationsbemühungen vor allem noch auf die Qualitätssicherung im Rahmen von beruflicher Aus- und Weiterbildung, hat Evaluation heutzutage in zahlreichen Forschungsvorhaben und Projekten der Berufsbildungsforschung und -zusammenarbeit Einzug gehalten. Evaluation als wesentlicher Bestandteil von Berufsbildungspolitik und -forschung wird eingesetzt, um Stärken und Ziele der Berufsbildung zu erfassen (Deutsche Bundesregierung 2019).

Obwohl konstatiert wird, dass Evaluationsvorhaben in der beruflichen Bildung über die letzten Jahre zugenommen haben, systematischer durchgeführt werden und immer mehr multiperspektivisch ausgerichtet sind, wird eine begrenzte Umsetzung der Evaluierungen auf Grundlage konkreter Standards als kritisch eingeschätzt (Beutner 2019, S. 121 f.).

Vor diesem Hintergrund untersucht der Beitrag die Frage, welche Standards für Evaluierungsvorhaben im Kontext der (internationalen) beruflichen Bildung genutzt werden können. Zunächst wird der Evaluationsbegriff im Kontext von Monitoring und Evaluationsforschung verortet, bevor auf die Standards der Deutschen Gesellschaft für Evaluation (DeGEval), der Organisation for Economic Cooperation and Development (OECD) sowie der United Nations (UN) eingegangen wird. In einem weiteren Schritt werden die drei Sets einer vergleichenden Würdigung unterzogen und es wird auf die Qualität von Evaluationen in der beruflichen Bildung eingegangen. Dabei wird auch gezeigt, inwiefern diese Standards eine Rolle für die Berufsbildungsforschung spielen, und ein abschließendes Fazit gezogen.

2 Evaluation im Kontext von Evaluationsforschung und Monitoring

Evaluationen werden in unterschiedlichsten Bereichen mit vielfältigsten Ansätzen und Methoden durchgeführt (Stockmann/Borrmann 2009, 263 f.). Der Versuch, Evaluation per se zu erfassen und einer Bestandsaufnahme zu unterziehen, scheint daher von vornherein zum Scheitern verurteilt und lässt sich allenfalls für bestimmte Politikfelder durchführen (Caspari 2009; Stockmann/Borrmann 2009, S. 264).

Wir verstehen Evaluation hier als Aktivitäten, um Ursachen und Wirkungen – auch in Relation zueinander – von Interventionen zu erfassen. Evaluation dient zudem der Erfassung bzw. der Einschätzung des (monetären und nicht-monetären) Wertes, des Nutzens und der Wirkung einer Intervention und ist daher üblicherweise periodisch angelegt, wie z. B. jährlich oder am Ende der Intervention. Eine Evaluation betrachtet nicht nur die Effizienz, sondern auch die Relevanz, Effektivität und Nachhaltigkeit von Interventionen. Evaluation liefert Nachweise dafür, weshalb Ziele und Ergebnisse (nicht oder nur teilweise) erreicht wurden, und kann dies idealerweise auch begründen (Scriven 1991, S. 139). Evaluation wird daher hier verstanden als „die mit politisch-administrativen Entscheidungs- und Handlungsprozessen verbundenen Analyse- und Erkenntnisverfahren [...], die darauf gerichtet sind, die Zielerreichung, die Wirkungen, Kosten und Nutzen politischer Entscheidungen, Programme und Maßnahmen zu identifizieren und in die Entscheidungs- und Handlungsprozesse ‚zurückzumelden‘ (feedback)“ (Wollmann 2017, S. 34). Evaluation kann zudem dazu dienen, summativ, d. h. bilanzierend und mit Blick auf die Ergebnisse einer Intervention, deren Wirkungen zu untersuchen sowie deren Ursachen zu analysieren (Stockmann 2021, S. 12).

In Abgrenzung dazu stellt die Evaluationsforschung eine Teilmenge aller Evaluationen dar, bei der empirische Evidenz systematisch gesammelt und auf Basis expliziter Kriterien bewertet wird (Treischl/Wolbring 2020, S. 19). Das systematische Vorgehen schließt den Einsatz von Methoden empirischer (Sozial-)Forschung ein, um Aussagen über die „(auch ursächliche) Wirksamkeit politischen und administrativen Handelns zu gewinnen“ (Wollmann 2017, S. 34). Evaluation und Evaluationsforschung haben sich parallel entwickelt und gelten deshalb als „siamesische Zwillinge“ (ebd.).

Im Kontext von Evaluation (und Evaluationsforschung) spielt Monitoring eine bedeutsame Rolle. Monitoring bezeichnet Aktivitäten mit dem Ziel, Informationen darüber zu sammeln, welchen Entwicklungsfortschritt eine Intervention zu einem bestimmten Zeitpunkt erreicht hat, und zwar in Relation zu den gesetzten Zielen und den geplanten Ergebnissen. Monitoring fokussiert also primär auf Effizienz und die Nutzung von Ressourcen. Sowohl Monitoring als auch Evaluation sind Planungsprozesse und schriftliche Produkte unterschiedlicher Arten von Interventionen und sind zugleich Teil der Qualitätssicherung. Sie weisen Überschneidungen auf, weil sie mitunter gleiche Ziele verfolgen und sich vergleichbarer Methoden bedienen (Markiewicz/Patrick 2016, S. 17 f.; UNICEF 2004, S. 108 ff.). Während Monitoring einen eher managementbezogenen Begriff im Sinne der Begleitung und Überprüfung eines laufenden Vorhabens darstellt, liegt der entscheidende Unterschied zur Evaluation in der Bewertungsfunktion, die mit Evaluation einhergeht (IFC/GTZ/DFID 2008, S. 22). Fasst man die einschlägige Literatur zu indikatorenbasiertem Monitoring und Evaluation (im Bildungsbereich) zusammen (z. B. Markiewicz/Patrick 2016, S. 4 ff.; OECD 2019, S. 5; Tilbury 2007, S. 242; Wollmann 2017, S. 41), lassen sich mehrere gemeinsame Ziele identifizieren:

- Systematische Generierung und Verbreitung von Erkenntnissen über Interventionen, z. B. durch Identifizierung von „best practice“ sowie hemmenden und fördernden Faktoren;
- Erfassung und Bewertung erwarteter und nicht erwarteter Ergebnisse von Interventionen;
- Überprüfung der Umsetzung und Erreichung von Interventionszielen auf Programm-, Institutionen- oder Politikstrategieebene, um bei Bedarf steuernd einzugreifen;
- Berichts- bzw. Rechenschaftslegung über Nutzung von Ressourcen, z. B. gegenüber Förderinstitutionen, politischen Akteuren und der (Fach-)Öffentlichkeit;
- Konstruktion einer basalen Grundlage für Entscheidungen, z. B., ob und inwiefern eine Intervention erfolgreich war oder ob eine Förderlinie mit Veränderungen weitergeführt werden sollte oder nicht.

Diese Konzepte bilden den Rahmen, in dem sich die folgende Auseinandersetzung mit Evaluationsstandards bewegt, die den Kern des Beitrags bilden.

3 Evaluationsstandards im Vergleich

3.1 Evaluationsstandards

Die Nützlichkeit und Akzeptanz einer Evaluation stehen und fallen mit ihrer Qualität. Zur Sicherung von Qualität werden wiederum Leitlinien, Richtlinien oder Standards benötigt, die zum einen als Orientierungs- und Bewertungsgrundlage für ein professionelles Handeln herangezogen werden können und zum anderen dem Schutz der Auftraggebenden sowie der Öffentlichkeit vor schädlichen Praktiken dienen sollen (Stockmann/Meyer 2014, S. 191; Treischl/Wolbring 2020, S. 24). Im Bildungsbereich existiert eine Vielfalt an verschiedenen Evaluationsstandards. Diese sind allerdings vorwiegend für den allgemeinbildenden Bereich (Müller/Renz 2019, S. 107) und den hochschulischen Bereich (Schmidt/Pohlenz 2019, S. 144 f.) vorgesehen, weniger für den berufsbildenden Bereich. Viele in (Bildungs-)Bereichen entwickelte Standards gehen im Kern auf übergeordnete und oft in der US-amerikanischen Tradition stehende Standards zurück (Beywl/Speer 2004, S. 69), von denen drei aufgrund der Reichweite ihrer Einsatzgebiete besonders markant erscheinen. Gemeint sind damit die Standards der DeGEval, der OECD-DAC (Development Assistance Committee) sowie der UN Evaluation Group (UNEG). Diese werden im Folgenden entlang ihres Entstehungskontextes, ihrer zentralen Merkmale sowie ihrer Einsatzmöglichkeiten vorgestellt und einer vergleichenden Würdigung unterzogen.

Standards für Evaluation der DeGEval

Die 1991 gegründete DeGEval hat als „sozialwissenschaftlich begründete Vorgehensweise“ (Beywl/Speer 2003, S. 13) „Standards für Evaluation“ formuliert, um „einen Orientierungsrahmen zur Bestimmung ‚guter‘ Evaluation [zu] geben“ (DeGEval 2016, S. 9) und somit die Qualität von Evaluationen zu sichern und zu entwickeln (ebd., S. 24). Für die Standards wird der Anspruch erhoben, politikfeldübergreifend einsetzbar zu sein, darunter explizit für die Bereiche Bildung, Entwicklungszusammenarbeit, Forschung und Innovation (ebd., S. 26). Evaluationen sollten vier grundlegende Eigenschaften aufweisen, nämlich (1) Nützlichkeit, (2) Durchführbarkeit, (3) Fairness und (4) Genauigkeit. Diese Eigenschaften werden auch als Standardgruppen bezeichnet und lassen sich wiederum in 25 Einzelstandards ausdifferenzieren. Beispielsweise gliedert sich die erste Standardgruppe „Nützlichkeit“, die sicherstellen soll, dass „die Evaluation sich an abgestimmten und geklärten [...] Evaluationszwecken sowie soweit möglich am Informationsbedarf der vorgesehenen [...] Nutzerinnen und Nutzer ausrichtet“ (ebd., S. 34), in acht Einzelstandards, darunter die Klärung der Evaluationszwecke, die Vollständigkeit und Klarheit der Berichterstattung sowie die Nutzung und der Nutzen der Evaluation.

Die Standards sind als Maximalstandards zu verstehen, die nicht vollumfänglich bei jeder Evaluation zum Einsatz kommen können, zumal einzelne Standards durchaus in einem „natürlichen Konkurrenzverhältnis“ (ebd., S. 28) zueinander zu sehen sind, so z. B. bei Genauigkeit versus Durchführbarkeit. Sie stehen in der Tradition der Methoden empirischer Sozialforschung und zu allen Standards existieren konkrete

Umsetzungshinweise, die wiederum weitgehend aus den JCSEE-Standards (Joint Committee on Standards for Educational Evaluation 1994; siehe auch Yarbrough/Shulha/Hopson/Caruthers 2011) abgeleitet sind. Wie breit die Standards aufgestellt sind, zeigt das Beispiel „Effizienz“, für das es heißt, „der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen. Begründung: Um die erforderliche Akzeptanz für Evaluation nicht zu gefährden, sollte immer das Aufwand-Nutzen-Verhältnis reflektiert werden. Idealerweise sollte also der [...] Nutzen einer Evaluation Kosten und Aufwand übertreffen oder wenigstens ausgleichen“ (DeGEval 2016, S. 40). In den Standards selbst wird auf die Schwierigkeit verwiesen, genaue Aussagen zu Aufwand und Nutzen einer Evaluation zu erheben, wobei solche Schwierigkeiten nicht dazu führen dürfen, dass auf die Erfassung der Effizienz verzichtet wird. Vielmehr soll bereits in der Planung berücksichtigt werden, welche Art von Effizienz erfasst werden soll und inwiefern der Aufwand in Relation zum (anti-zipierten) Nutzen steht.

Qualitätsstandards für die Entwicklungsvaluierung der OECD-DAC

Die DAC-Qualitätsstandards für die Entwicklungsvaluierung wurden 2006 in Form einer vorläufigen Fassung genehmigt und 2010 offiziell angenommen und veröffentlicht (OCED 2010, S. 1). Im Kern zielen sie darauf, die Qualität von Evaluierungen zu steigern und dadurch den potenziellen Beitrag zu stärken, den eine Evaluierung zur Verbesserung der Entwicklungsergebnisse leisten kann. Sie wurden in erster Linie für die DAC-Mitglieder aufgestellt und bilden einen Leitfaden für „gute Praxis“, wobei ihre Anwendung nicht verpflichtend ist (ebd., S. 5). Die Standards sind in Kopplung mit den DAC-Grundsätzen für die Evaluierung der Entwicklungszusammenarbeit¹ zu betrachten. Während die Grundsätze auf das Management und den institutionellen Aufbau von Evaluierungssystemen fokussieren, fungieren die vorliegenden Standards hingegen als Anleitung für Evaluierungsprozesse und -produkte (ebd.).

Wie bereits impliziert, beziehen sich die Standards vor allem auf die Entwicklungsvaluierung, unter der „die systematische und objektive Bewertung einer noch laufenden bzw. bereits abgeschlossenen Entwicklungsmaßnahme, ihrer Konzeption, Durchführung und Ergebnisse“ (ebd., S. 6) verstanden wird. Die Struktur der Standards reflektiert den üblichen Ablauf eines Evaluierungsprozesses (ebd., S. 5). Das Standardset gliedert sich dabei in vier Sektionen, nämlich 1) Allgemeine Betrachtungen, 2) Zweck, Planung und Konzeption, 3) Durchführung und Berichterstattung sowie 4) Follow-up, Nutzung der Evaluierung, Lernprozess. Insgesamt umfasst das Set 37 Einzelstandards, die in konkreten Evaluierungsvorhaben entsprechend nationaler Kontexte sowie Ziele der Evaluierungen anzupassen sind (ebd.).

¹ Die OECD-DAC hat 1991 „Principles for Evaluation of Development Assistance“ dargelegt, welche aus Sicht der DAC-Mitglieder die wichtigsten Anforderungen an einen Evaluationsprozess widerspiegeln. Die Prinzipien wurden für die Evaluierung von laufenden und abgeschlossenen Vorhaben der Entwicklungshilfe aufgestellt. Jedes Prinzip enthält dabei Empfehlungen hinsichtlich der Nutzung von Evaluationen für bestimmte Entwicklungshilfemaßnahmen (OECD 1991, S. 4).

Die Normen und Standards der UNEG

Ein drittes Beispiel sind die „Norms and Standards for Evaluation“ der UNEG (2016). Diese wurden 2005 erstmalig veröffentlicht und dienen als Rahmen für Evaluationskompetenzen, Peer Reviews und Benchmarkings (UNEG 2016, S. 8). Grundsätzlich wird in dem UNEG-Dokument zwischen Normen und Standards differenziert. Zunächst können zwei Gruppen von Normen unterschieden werden, nämlich zehn allgemeine Normen (z. B. Nützlichkeit, Glaubwürdigkeit, Transparenz) und vier spezifische Normen, die sich nur auf die innerinstitutionelle Anwendung bei UN-Einrichtungen beziehen (ebd., S. 10 ff.). Neben den Normen existieren fünf Standards bzw. Standardgruppen für Evaluation (Institutioneller Rahmen, Management der Evaluationsfunktion, Evaluationskompetenz, Evaluationsdurchführung, Qualität), die jeweils wiederum in insgesamt 24 Standards untergliedert sind. Darunter befinden sich Standards zur Leitung einer Evaluation, ethische Aspekte der Evaluationskompetenz oder methodische Standards für die Durchführung von Evaluationen (ebd., S. 7).

Bei den Standards der UNEG steht als übergeordnetes Ziel von Evaluationen die Rechenschaft und das Lernpotenzial, d. h. das Verständnis von erwarteten und unerwarteten Ergebnissen von Interventionen, im Vordergrund; erst daraus werden weitere Ziele für die Planung, die Budgetierung oder die Implementation abgeleitet (ebd., S. 10)

3.2 Vergleichende Würdigung der Standards

Ein Vergleich der Standards kann angesichts des Beitragsumfangs nur vor dem Hintergrund ausgewählter Aspekte in groben Zügen vorgenommen werden (vgl. Abb. 1).

Eine Gemeinsamkeit, welche die vorgestellten drei Sets an Normen und Standards teilen, ist ihr hochgradig normativer Charakter. Im Kern zielen sie darauf, vorzugeben bzw. zu empfehlen, wie eine Evaluation zu erfolgen hat, damit sie als eine „gute Praxis“ bewertet werden kann. Darüber hinaus haben alle drei die klare Intention zum Ziel, die Ergebnisse der Evaluation dafür zu nutzen, Rückschlüsse für informierte Entscheidungen zu treffen (UNEG 2016, S. 10).

Gleichwohl die Standards mit Blick auf ihre Reichweite und ihren Detailgrad voneinander abweichen, sind multiple inhaltliche Überschneidungen feststellbar. Beispielsweise sind in allen dreien Standards dargelegt, welche um einen Prozess der Planung, Durchführung und Nachbereitung einer Evaluation herum strukturiert sind. Bei der DeGEval (2016, S. 44 ff.) sind diese unter der Standardgruppe „Genauigkeit“ aufgeführt, bei der UNEG (2016, S. 21 ff.) unter „Conduct of Evaluations“. Bei der OECD (2010, S. 5) entspricht hingegen der gesamte Aufbau des Dokuments der Logik eines typischen Evaluierungsprozesses. Daran anschließend lassen sich in den Dokumenten auch Empfehlungen für eine vollständige und klare Berichterstattung und eine systematische Dissemination der Evaluationsergebnisse finden (DeGEval 2016, S. 37 f.; OECD 2010, S. 15; UNEG 2016, S. 26 f.). In allen drei Dokumenten sind auch Anforderungen an Qualifikationen und Kompetenzen der Evaluierenden formuliert. In den DeGEval-Standards wird festgehalten, dass Evaluierende „fachlich und methodisch kompetent sein [sollen], damit für die Evaluation und ihre Ergebnisse ein

Titel	Standards für Evaluation	Qualitätsstandards für die Entwicklungsevaluierung	Norms and Standards for Evaluation
Akteur	DeGEval	OECD-DAC	UNEG
Erstfassung	2002	2006, 2010	2005
Standards	Standardgruppen und Standards	Qualitätsstandards	Norms and Standards
	<p>Nützlichkeit</p> <ul style="list-style-type: none"> Identifizierung der Beteiligten und Betroffenen Klärung der Evaluationszwecke Kompetenz und Glaubwürdigkeit des Evaluators/der Evaluatorin Auswahl und Umfang der Informationen Transparenz von Werthaltungen Vollständigkeit und Klarheit der Berichterstattung Rechtzeitigkeit der Evaluation Nutzung und Nutzen der Evaluation <p>Durchführbarkeit</p> <ul style="list-style-type: none"> Angemessene Verfahren Diplomatisches Vorgehen Effizienz von Evaluation <p>Fairness</p> <ul style="list-style-type: none"> Formale Vereinbarungen Schutz individueller Rechte Umfassende und faire Prüfung Unparteiliche Durchführung und Berichterstattung Offenlegung von Ergebnissen und Berichten <p>Genauigkeit</p> <ul style="list-style-type: none"> Beschreibung des Evaluationsgegenstandes Kontextanalyse Beschreibung von Zwecken und Vorgehen Angabe von Informationsquellen Valide und reliable Informationen Systematische Fehlerprüfung Angemessene Analyse qualitativer und quantitativer Informationen Begründete Bewertungen und Schlussfolgerungen Metaevaluation 	<p>Allgemeine Betrachtungen</p> <ul style="list-style-type: none"> Entwicklungsevaluierung Freier und offener Evaluierungsprozess Evaluierungsethik Partnerschaftlicher Ansatz Koordinierung und Partnerausrichtung Kapazitätsentwicklung Qualitätskontrolle <p>Zweck, Planung und Konzeption</p> <ul style="list-style-type: none"> Anlass und Zweck der Evaluierung Spezifische Ziele der Evaluierung Gegenstand und Reichweite der Evaluierung Evaluierbarkeit Einbeziehung der betroffenen Akteure Systematische Untersuchung der Möglichkeit einer Gemeinschaftsevaluierung Evaluierungsfragen Auswahl und Anwendung der Evaluierungskriterien Auswahl von Ansatz und Methoden Ressourcen Governance- und Managementstrukturen Leistungsbeschreibung <p>Durchführung und Berichterstattung</p> <ul style="list-style-type: none"> Evaluierungsteam Unabhängigkeit der Evaluatoreninnen oder Evaluatoren von den betroffenen Akteuren Konsultation und Schutz der betroffenen Akteure Fristgerechte und budgetkonforme Durchführung der Evaluierung Evaluierungsbericht Klarheit und Repräsentativität der Zusammenfassung Kontext der Entwicklungsmaßnahme Wirkungslogik Validität und Verlässlichkeit der Informationsquellen Erklärung der verwendeten Methoden Klarheit der Analyse Beantwortung der Evaluierungsfragen Hinweis auf Veränderungen und Einschränkungen der Evaluierung Hinweis auf Unstimmigkeiten zwischen den Mitgliedern des Evaluierungsteams Aufnahme von Kommentaren betroffener Akteure <p>Follow-up, Nutzung der Evaluierung, Lernprozess</p> <ul style="list-style-type: none"> Aktualität, Relevanz und Nutzung der Evaluierung Systematische Beantwortung und Weiterverfolgung der Empfehlungen Verbreitung der Ergebnisse 	<p>General Norms for Evaluation</p> <ul style="list-style-type: none"> Internationally agreed principles, goals and targets Utility Credibility Independence Impartiality Ethics Transparency Human rights and gender equality National evaluation capacities Professionalism <p>Institutional Norms for Evaluation in the United Nations Systems</p> <ul style="list-style-type: none"> Enabling environment Evaluation policy Responsibility for the evaluation function Evaluation use and follow-up <p>Standards for Evaluation</p> <p>Institutional Framework</p> <ul style="list-style-type: none"> Institutional framework for evaluation Evaluation policy Evaluation plan and reporting Management response and follow-up Disclosure policy <p>Management of the Evaluation Function</p> <ul style="list-style-type: none"> Head of evaluation Evaluation guidelines Responsiveness of the evaluation function <p>Evaluation Competencies</p> <ul style="list-style-type: none"> Competencies Ethics <p>Conduct of Evaluations</p> <ul style="list-style-type: none"> Timeliness and intentionality Evaluability assessment Terms of reference Evaluation scope and objectives Methodology Stakeholder engagement and reference groups Human rights-based approach and gender mainstreaming strategy Selection and composition of evaluation terms Evaluation report and products Recommendations Communication and dissemination <p>Quality</p> <ul style="list-style-type: none"> Quality assurance system Quality control of the evaluation design Quality control at the final stage of evaluation

Abbildung 1: Übersicht zentraler Evaluationsnormen und -standards der DeGEval, OECD-DAC und UNEG (Zusammenstellung der Autor*innen)

Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird“ (DeGEval 2016, S. 36). Die Standards der UNEG determinieren sogar, dass all jene, die am Design, an der Durchführung und am Management von Evaluationen beteiligt sind, über entsprechende „core competencies“ verfügen sollten. Diese beziehen sich auf adäquate Qualifikationen, Fertigkeiten, Eigenschaften und Erfahrungen, mit denen die Glaubwürdigkeit und die Qualität des Evaluationsprozesses gewährleistet werden soll (UNEG 2016, S.19). In den DAC-Qualitätsstandards werden an diesem Punkt Spezifika der Entwicklungsarbeit sichtbar. So wird gefordert, dass Evaluierende sowohl über einschlägige Qualifikationen im Bereich der Evaluation als auch über fachliche Kenntnisse zur jeweiligen Thematik verfügen sollen. Ferner ist bei der Zusammensetzung von Evaluierungsteams auf ein ausgewogenes Geschlechterverhältnis zu achten und es sind Fachleute aus den betroffenen Partnerländern bzw. -regionen einzubeziehen

(OECD 2010, S. 11). Auch weitere Standards, wie etwa die Glaubwürdigkeit, die Nützlichkeit, die Unabhängigkeit und die Fairness von Evaluationen, sind in allen drei Sets zu finden (z. B. UNEG 2016, S. 10 f.).

Es existieren allerdings auch Unterschiede, die sich vor allem aus dem Entstehungskontext der Standards ergeben. Die Normen und Standards der UNEG wurden vor allem zur Stärkung und Harmonisierung der Evaluationsvorhaben des UN-Systems entwickelt. Diese sollten einen Rahmen für UNEG-Evaluierungskompetenzen, Peer Reviews und Benchmarkings bilden (ebd., S. 8). Sowohl in den Normen als auch in den Standards wird die Integration von menschenrechtlichen und geschlechtergleichstellenden Aspekten in alle Phasen der Evaluation ausdrücklich gefordert (ebd., S. 12 und 24 f.). Demgegenüber wurden die DAC-Qualitätsstandards primär für die Evaluierung von laufenden und abgeschlossenen Vorhaben der Entwicklungsarbeit aufgestellt, um die Qualität von Entwicklungsprozessen und -produkten zu erhöhen (OECD 2010, S. 5 f.). In diesem Kontext spiegeln sich auch Besonderheiten der Entwicklungsarbeit in den Standards wider, wie etwa in den Empfehlungen zur Verfolgung eines partnerschaftlichen Ansatzes oder zur Unterstützung der Kapazitätsentwicklung der Entwicklungspartner (ebd., S. 7). Qualität, genauer gesagt Qualitätssicherung und -entwicklung, stellt auch für die DeGEval (2016, S. 24) das übergreifende Ziel dar, das mit der Entwicklung der Standards anvisiert wurde. Allerdings fokussieren die DeGEval-Standards nicht nur auf die Entwicklungszusammenarbeit, sondern es wird für sie der Anspruch erhoben, politikfeldübergreifend zu gelten, so auch für den Bildungsbereich, das Gesundheitswesen oder die soziale Arbeit (ebd., S. 26). Qualitätsmanagement wird von der UNEG unter dem Standard „Quality“ erfasst. Hier werden mitunter die Bereitstellung eines Qualitätssicherungssystems sowie die Qualitätskontrolle in der Design- und Abschlussphase der Evaluation dezidiert empfohlen (UNEG 2016, S. 28 f.).

4 Evaluationsstandards in der beruflichen Bildung

Inwiefern welche Evaluationsstandards in welchem Ausmaß in der Praxis und/oder Forschung mit Bezug zur beruflichen Bildung Anwendung finden (können), ist kaum bekannt. Lediglich für die DeGEval-Standards sind Beywl und Speer der Frage nachgegangen, ob „die auf breite Anwendungsfelder gerichteten, auf die nordamerikanische Tradition aufbauenden DeGEval-Standards auch in institutionell und curricular unterschiedlich geprägten Berufsbildungssystemen Geltung beanspruchen [können]“ (Beywl/Speer 2003, S. 13). Hierbei hat sich gezeigt, dass zwar nicht jeder einzelne Standard in jedem Evaluierungsprojekt nutzbar ist, dass aber die „Anwendbarkeit jedes einzelnen DeGEval-Standards in unterschiedlichen Kontexten der Berufsbildung bestätigt“ werden kann (ebd., S. 14). Darüber hinaus ist in der deutschsprachigen Literatur zur (internationalen) Berufsbildungsforschung wenig über die Qualität von Evaluationen bekannt. Dort, wo Berufsbildungsforschung an internationale (Berufsbildungs-) Zusammenarbeit grenzt, existiert eine vergleichsweise überschaubare Anzahl an Untersuchungen, die vorwiegend von wenigen Forschungsteams durchgeführt werden, so

etwa von einem Saarbrücker Forschungsteam, das innerhalb von knapp 30 Jahren auf 83 Analysen bilateraler Berufsbildungsmaßnahmen kommt (Stockmann 2021, S. 17). Weitere Untersuchungen über die Qualität von Evaluationen ergaben teilweise „erschreckende Ergebnisse“ (Caspari 2009, S. 186). Entlang von Metaevaluationen zur Qualität von Evaluationen wird aufgezeigt, dass Datenerhebungen weitgehend unsystematisch und inkorrekt durchgeführt wurden. Mit Verweis auf eine Studie von Forss und Carlsson (1997, S. 488) wird resümiert, dass „insgesamt [...] bei dem Großteil der Evaluationen keine Daten im sozialwissenschaftlichen Sinne erhoben [wurden], die die Evaluator(inn)en zu ihren Schlussfolgerungen hätten führen können. Es sind praktisch keine Daten vorhanden, so dass der Evaluationsprozess mit wenigen Ausnahmen als nicht transparent gewertet werden muss“ (Caspari 2009, S. 187).

Vor diesem Hintergrund kann den hier vorgestellten Sets an Evaluationsstandards attestiert werden, dass sie zwar hilfreich und für Evaluationen notwendig, aber nicht hinreichend sind: Die vorgestellten Standards haben keinen verpflichtenden Charakter, weswegen sie nicht automatisch als Garant für eine hohe Qualität von Evaluationen betrachtet werden können (Stockmann/Meyer 2014, S. 195). Nichtsdestotrotz können sie in einem Evaluierungsprozess bei Bestrebungen der Qualitätssicherung und -entwicklung Orientierung bieten, indem sie im Zusammenspiel mit Evaluationskriterien, nationalen Richtlinien und konkreten Evaluationsfragen einen übergeordneten Rahmen bilden (OECD 2021, S. 21).

Anhand des oben skizzierten Vergleichs zeigt sich, dass alle drei Standardsets in mehreren Kontexten verortet sind. Demzufolge können sie für eine Anwendung in einem spezifischen Evaluationsvorhaben nicht pauschal übernommen werden, sondern sie sind an gegebene Rahmenbedingungen und mit Blick auf Evaluationsgegenstände anschlussfähig zu machen. Beispielsweise könnten die DeGEval-Standards theoretisch für ein breites Spektrum an nationalen und internationalen Berufsbildungsforschungsvorhaben angewendet werden. Hingegen könnten für Vorhaben mit Schwerpunkt auf internationaler Berufsbildungszusammenarbeit die Standards der OECD-DAC und der UNEG fruchtbare Anknüpfungsmöglichkeiten bieten.

5 Fazit

In unserem Beitrag haben wir drei Sets an Standards vorgestellt, die unterschiedlichen Kontexten entspringen, aber alle (auch) für die Berufsbildungsforschung und vor allem für die internationale Berufsbildungszusammenarbeit genutzt werden können. Während es unwahrscheinlich ist, dass alle Standards gleichermaßen in einer Evaluation Anwendung finden, bietet der Vergleich die Möglichkeit, die unterschiedlichen Kontexte und die differierende Reichweite der Standards für einen konkreten Einsatz zu überprüfen. Obwohl die vorgestellten Standards zweifelsohne einen wesentlichen Meilenstein in der Professionalisierung der Evaluation und der Evaluierenden markieren (Beywl/Böttcher/Hense/Widmer 2019, S. 15), bleibt trotzdem noch weitestgehend offen, inwiefern die Standards in welchen Kontexten innerhalb der Be-

rufsbildungsforschung genau angewendet werden. Dieses Desiderat ist in weiteren Untersuchungen zu erörtern, bedenkt man, dass Evaluation in jedem Fall maßgeblich der Transparenz und Qualitätssicherung von Interventionen dient, sei dies in der Berufsbildungsforschung oder jenseits dieses Feldes.

Literaturverzeichnis

- Beutner, Marc (2019): Die DeGEval-Standards in der beruflichen Bildung. In: Hense, Jan et al. (Hrsg.): Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis? Münster, S. 121–136.
- Beywl, Wolfgang/Böttcher, Wolfgang/Hense, Jan/Widmer, Thomas (2019): Vorläufer, Entstehung und Weiterentwicklung der Standards für Evaluation. In: Hense, Jan et al. (Hrsg.): Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis? Münster, S. 15–32.
- Beywl, Wolfgang/Speer, Sandra (2003): Was kennzeichnet gute Evaluationen in der Berufsbildung? In: Berufsbildung in Wissenschaft und Praxis, 32 (20155), S. 13–17.
- Beywl, Wolfgang/Speer, Sandra (2004): Developing standards to evaluate vocational education and training. In: Descy, Pascaline/Tessaring, Manfred (Hrsg.): The foundations of evaluation and impact research. Third report on vocational training research in Europe: Background report. Luxembourg, S. 51–129.
- Brosi, Walter (2003): Bedeutung der Evaluationsforschung nimmt zu. In: Berufsbildung in Wissenschaft und Praxis, 32 (20155), S. 3–4.
- Caspari, Alexandra (2009): ‚Rigorese‘ Wirkungsevaluation – methodische und konzeptionelle Ansätze der Wirkungsmessung in der Entwicklungszusammenarbeit. In: Zeitschrift für Evaluationsforschung, 8(2), S. 183–213.
- DeGEval (2016): Standards für Evaluation. Online: https://www.degeval.org/fileadmin/Publikationen/DeGEval-Standards_fuer_Evaluation.pdf (20.07.2021).
- Deutsche Bundesregierung (2019): Strategie der Bundesregierung zur internationalen Berufsbildungszusammenarbeit. Online: https://www.bmbf.de/files/137_19_Strategie_Bundesregierung.pdf (02.02.2021).
- Forss, Kim/Carlsson, Jerker (1997): The quest for quality – or can evaluation findings be trusted? In: Evaluation, (3/4), S. 481–501.
- Gonon, Philipp (2008): Qualitätssicherung in der beruflichen Bildung. Eine Bilanz. In: Klieme, Eckhard/Tippelt, Rudolf (Hrsg.): Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz. Zeitschrift für Pädagogik. 53. Beiheft. Weinheim, Basel, S. 96–107.
- IFC/GTZ/DFID (2008): The Monitoring and Evaluation Handbook for Business Environment Reform. A Handbook for Practitioners. Danvers, MA.
- Joint Committee on Standards for Educational Evaluation (1994): The program evaluation standards. How to assess evaluations of educational programs. 2. Auflage. Thousand Oaks, CA.

- Marchart, Oliver (2013): *Das unmögliche Objekt. Eine postfundamentalistische Theorie der Gesellschaft*. Frankfurt a. M.
- Markiewicz, Anne/Patrick, Ian (2016): *Developing Monitoring and Evaluation Frameworks*. Los Angeles, London.
- Müller, Sabine/Renz, Monika (2019): Evaluationsstandards im Handlungsfeld Schule. In: Hense, Jan et al. (Hrsg.): *Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?* Münster, S. 107–120.
- OECD (1991): *Principles for Evaluation of Development Assistance*. Online: <https://www.oecd.org/dac/evaluation/2755284.pdf>. (30.06.2021).
- OECD (2010): *Qualitätsstandards für die Entwicklungsevaluierung*. Online: <https://doi.org/10.1787/9789264085183-de> (25.11.2022).
- OECD (2019): *Better Criteria for Better Evaluation. Revised Evaluation Criteria. Definitions and Principles for Use*. Online: <https://www.oecd.org/dac/evaluation/revised-evaluation-criteria-dec-2019.pdf> (04.06.2021).
- OECD (2021): *Applying Evaluation Criteria Thoughtfully*. Paris.
- Schmidt, Uwe/Pohlenz, Philipp (2019): Evaluationsstandards im Handlungsfeld Hochschule. In: Hense, Jan et al. (Hrsg.): *Evaluation: Standards in unterschiedlichen Handlungsfeldern. Einheitliche Qualitätsansprüche trotz heterogener Praxis?* Münster, S. 137–150.
- Scriven, Michael (1991): *Evaluation Thesaurus*. 4. Auflage. Thousand Oaks.
- Stockmann, Reinhard (2021): *Systematic Evaluation Analysis als Instrument der Wirkungsevaluation*. In: Stockmann, Reinhard/Ertl, Hubert (Hrsg.): *Evaluation und Wirkungsforschung in der beruflichen Bildung*. Bonn, S. 11–24.
- Stockmann, Reinhard/Borrmann, Axel (2009): *Das Evaluationssystem in der deutschen Entwicklungszusammenarbeit*. In: *Zeitschrift für Evaluation*, 8(2), S. 263–295.
- Stockmann, Reinhard/Meyer, Wolfgang (2014): *Evaluation. Eine Einführung*. 2. Auflage. Opladen, Bloomfield Hills.
- Tilbury, Daniella (2007): *Monitoring and Evaluation during the UN Decade of Education for Sustainable Development*. In: *Journal of Education for Sustainable Development*, 1(2), S. 239–254.
- Treischl, Edgar/Wolbring, Tobias (2020): *Wirkungsevaluation. Grundlagen, Standards, Beispiele*. Weinheim, Basel.
- UNEG (2016): *Norms and Standards for Evaluation*. New York.
- UNICEF (2004): *Programme Policy and Procedure Manual: Programme Operations*. New York.
- Wollmann, Hellmut (2017): *Entwicklungslinien von Evaluation und Evaluationsforschung in Deutschland. Zwischen Vorgeschichte, Aufbruch und Konsolidierung*. In: *Zeitschrift für Evaluation*, 16(2), S. 33–53.
- Yarbrough, Donald B./Shulha, Lyn M./Hopson, Rodney K./Caruthers, Flora A. (2011): *The program evaluation standards. A guide for evaluators and evaluation users*. 3. Auflage. Thousand Oaks, CA.

Autorin und Autor

Prof.in Dr.in Sandra Bohlinger

Institut für Berufspädagogik und Berufliche Didaktiken, Technische Universität
Dresden

Hoang Long Nguyen, M. A.

Wissenschaftlicher Mitarbeiter Institut für Berufspädagogik und Berufliche Didaktiken, Technische Universität Dresden