FAU Studien aus der Philosophischen Fakultät 5

Mira Beck

Aus Evaluationen lernen: Durchführung einer Meta-Evaluation zur Generierung von Empfehlungen für die Evaluationspraxis

FAU UNIVERSITY PRESS

Aus Evaluationen lernen: Durchführung einer Meta-Evaluation zur Generierung von Empfehlungen für die Evaluationspraxis

FAU Studien aus der Philosophischen Fakultät

Band 5

Herausgeber der Reihe:
Prof. Dr. Michele Ferrari und Prof. Dr. Rainer Trinczek

Mira Beck

Aus Evaluationen lernen: Durchführung einer Meta-Evaluation zur Generierung von Empfehlungen für die Evaluationspraxis

Erlangen
FAU University Press
2016

Bibliografische Information der Deutschen Nationalbibliothek:
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.d-nb.de abrufbar.

Das Werk, einschließlich seiner Teile, ist urheberrechtlich geschützt.
Die Rechte an allen Inhalten liegen bei ihren jeweiligen Autoren.
Sie sind nutzbar unter der Creative Commons Lizenz BY-NC-ND.

Der vollständige Inhalt des Buchs ist als PDF über den OPUS Server der Friedrich-Alexander-Universität Erlangen-Nürnberg abrufbar:
https://opus4.kobv.de/opus4-fau/home

Umschlagfoto: www.shutterstock.com/g/donskarpo

Verlag und Auslieferung:
FAU University Press, Universitätsstraße 4, 91054 Erlangen

Druck: docupoint GmbH

ISBN: 978-3-944057-67-5 (Druckausgabe)
eISBN: 978-3-944057-68-2 (Online-Ausgabe)
ISSN: 2363-720X

Aus Evaluationen lernen: Durchführung einer Meta-Evaluation zur Generierung von Empfehlungen für die Evaluationspraxis

Der Philosophischen Fakultät

der Friedrich-Alexander-Universität

Erlangen-Nürnberg

zur

Erlangung des Doktorgrades Dr. phil.

vorgelegt von

Mira Beck

aus

Nürnberg

Als Dissertation genehmigt

von der Philosophischen Fakultät

der Friedrich-Alexander-Universität Erlangen-Nürnberg

Tag der mündlichen Prüfung: 20.06.2016

Vorsitzende des Promotionsorgans: Prof. Dr. Heike Paul

Gutachter: Prof. Dr. Mark Stemmler

Gutachterin: Prof. Dr. Sabine Walper

Inhaltsverzeichnis

Abkürzungsverzeichnis v
Danksagung vii
Zusammenfassung ix
Abstract xi
1 Einleitung 1
2 Theoretischer Hintergrund 5
2.1 Evaluation 5
2.1.1 Definition und Begriffsklärung 5
2.1.2 Entwicklung der Evaluationsforschung 6
2.1.3 Charakteristika zur Einordnung von Evaluationen 8
2.1.3.1 Anwendungsbereiche und Gegenstände von Evaluationen 8
2.1.3.2 Ziele und Funktionen von Evaluationen 9
2.1.3.3 Zeitpunkt der Evaluation 12
2.1.3.4 Interne und externe Evaluationen 13
2.1.3.5 Evaluationsansätze 15
2.1.4 Phasen eines Evaluationsprozesses 16
2.1.5 Bedeutung und Grenzen von Evaluationen 24
2.1.6 Nutzen und Nutzung von Evaluationen 26
2.1.6.1 Nutzenerwartungen 26
2.1.6.2 Nutzung von Evaluationsergebnissen 27
2.1.6.3 Einflüsse auf die Nutzung 28
2.1.7 Evaluation zwischen Wissenschaft und Praxis 30
2.1.7.1 Evaluationsforschung und Grundlagenforschung 30
2.1.7.2 Evaluation als sozialer und politischer Prozess 33
2.1.7.3 Berufsbild von Evaluatoren 35

2.2 Evaluationsstandards und Meta-Evaluation ... 38
2.2.1 Evaluationsstandards – Einführung ... 38
2.2.2 Standards und Richtlinien für Evaluationen – ein Überblick ... 40
2.2.3 Die Standards für Evaluation der DeGEval ... 43
2.2.3.1 Nützlichkeitsstandards ... 44
2.2.3.2 Durchführbarkeitsstandards ... 45
2.2.3.3 Fairnessstandards ... 46
2.2.3.4 Genauigkeitsstandards ... 47
2.2.3.5 Hinweise zu den Standards und ihrer Anwendung ... 48
2.2.4 Meta-Evaluation – Begriffsklärung und Einführung ... 49
2.2.5 Forschungsstand bezüglich Meta-Evaluationen und Evaluationsstandards ... 54
2.2.5.1 Methodik bei Meta-Evaluationen: Generelles Vorgehen ... 55
2.2.5.2 Methodik bei Meta-Evaluationen: Anwendung von Evaluationsstandards ... 57
2.2.5.3 Befunde zu den amerikanischen Evaluationsstandards des JCSEE ... 61
2.2.5.4 Befunde zu den deutschen Standards für Evaluation ... 65
2.2.5.5 Zusammenfassung des Forschungsstandes ... 70
3 Ziele und Fragestellungen ... 73
4 Meta-Evaluation – Bewertung einer Evaluationsstudie ... 77
4.1 Die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ ... 77
4.1.1 Das Bundesprogramm „Elternchance ist Kinderchance“ ... 77
4.1.1.1 Qualifizierung von Fachkräften zu Elternbegleitern ... 78
4.1.1.2 Erprobung lokaler Vernetzungsstrukturen (Elternbegleitung Plus) ... 79
4.1.2 Rahmenbedingungen der Evaluation ... 80

4.1.3 Ziele, Design und Methoden der Evaluation 84
4.1.3.1 Evaluationsziele 84
4.1.3.2 Evaluationsdesign und Methodik 85
4.1.4 Zentrale Ergebnisse 98
4.1.4.1 Zentrale Ergebnisse aus Modul 1 98
4.1.4.2 Zentrale Ergebnisse aus Modul 2 103
4.1.4.3 Zentrale Ergebnisse aus Modul 3 107
4.2 Bewertung der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“: Die Meta-Evaluation 111
4.2.1 Methode der Meta-Evaluation 111
4.2.2 Bewertung der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ anhand von Evaluationsstandards 120
4.2.2.1 Nützlichkeitsstandards 121
4.2.2.2 Durchführbarkeitsstandards 153
4.2.2.3 Fairnessstandards 164
4.2.2.4 Genauigkeitsstandards 178
4.2.3 Zusammenfassung der Ergebnisse der Meta-Evaluation 198
5 Diskussion 201
5.1 Zentrale Herausforderungen im Evaluationsprozess 201
5.1.1 Besonderheiten des Evaluandums 202
5.1.2 Planung und Durchführung der Evaluation 203
5.1.3 Berücksichtigung der Stakeholder 208
5.1.3.1 Beteiligung der Stakeholder bei der Evaluationsplanung und -durchführung 209
5.1.3.2 Passende Berichtsformen 211
5.1.3.3 Information über Ergebnisse 212
5.1.3.4 Unterstützung bei der Ergebnisnutzung 213
5.1.3.5 Zusammenfassende Hinweise 214

5.1.4 Glaubwürdigkeit und Diplomatie 214
5.1.5 Genauigkeit vs. Durchführbarkeit 217
5.2 Bewertung der DeGEval-Standards 218
5.2.1 Die DeGEval-Standards als Bewertungsinstrument bei einer Meta-Evaluation 218
5.2.2 Zur Verwendung der DeGEval-Standards bei der Durchführung von Evaluationen 222
5.2.3 Zur Diskussion über allgemeine vs. Feldspezifische Evaluationsstandards 225
5.2.4 Vorschläge für eine Überarbeitung der DeGEval-Standards 226
5.3 Kritische Reflexion der angewandten Methode 230
5.3.1 Stärken und Schwächen der vorliegenden Meta-Evaluation . 231
5.3.2 Geltungsbereich 237
5.3.3 Weitere methodische Überlegungen 239
5.4 Fazit zur vorliegenden Studie und Ausblick 242
6 Literaturverzeichnis 245
7 Tabellen- und Abbildungsverzeichnis 257
8 Anhang 259

Abkürzungsverzeichnis

BMFSFJ	Bundesministerium für Familie, Senioren, Frauen und Jugend
DeGEval	Gesellschaft für Evaluation e. V.
DJI	Deutsches Jugendinstitut e. V.
FAU	Friedrich-Alexander-Universität Erlangen-Nürnberg
JCSEE	Joint Committee on Standards for Educational Evaluation

Danksagung

Bei der Erstellung dieser Dissertation habe ich von vielen Menschen Unterstützung erfahren, für die ich mich an dieser Stelle bedanken möchte.

Zunächst danke ich meinem Doktorvater Prof. Dr. Mark Stemmler für die Möglichkeit, bei ihm zu promovieren. Ich habe das Doktorat als lehrreiche und herausfordernde Phase erlebt und bedanke mich für das Vertrauen, das mein Betreuer mir entgegenbrachte. Mein Dank gilt auch meiner Zweitgutachterin Prof. Dr. Sabine Walper sowie meinem Drittprüfer Prof. Dr. Rainer Trinczek.

Die vorliegende Arbeit entstand im Rahmen der im Auftrag des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) durch das Deutsche Jugendinstitut e. V. (DJI) und die Friedrich-Alexander-Universität durchgeführten Evaluation des Bundesprogramms „Elternchance ist Kinderchance“. Ich möchte mich bei allen am Programm und an der Evaluation beteiligten Personen bedanken, ohne die diese Arbeit letztlich nicht möglich gewesen wäre. Mein besonderer Dank gilt denjenigen, die durch ihre Stellungnahmen maßgeblich zum Gelingen meiner Studie beigetragen haben. Insbesondere Dagmar Müller vom DJI sowie Ulrich Paschold und Christine Mühlbach vom BMFSFJ haben mich mit ihren hilfreichen Anmerkungen sehr unterstützt, wofür ihnen herzlich gedankt sei.

Bei meinen Kolleginnen und Kollegen möchte ich mich für zahlreiche Ratschläge und auch für persönliche Unterstützung bedanken. Hervorheben möchte ich Dr. Doris Bender, die mir stets ein offenes Ohr bot. Selbiges gilt für M. Sc. Mariann Schwaß, die mich darüber hinaus auch durch viele fachliche Diskussionen und Anregungen unterstützte, wofür ich ihr ganz herzlich danke.

Zuletzt und am meisten möchte ich meiner Familie, meinem Partner und meinen Freunden danken. Sie haben stets an mich geglaubt, mich ermutigt und mir Verständnis entgegengebracht. Für diese Unterstützung danke ich ihnen von ganzem Herzen.

Erlangen, Januar 2016 *Mira Beck*

Zusammenfassung

Evaluationsstandards und Meta-Evaluationen (im Sinne der Evaluation von Evaluationsstudien) stellen wichtige Maßnahmen zur Qualitätssicherung von Evaluationen dar. Umso gravierender ist zu bewerten, dass bislang ein Mangel an Forschungsarbeiten zu Meta-Evaluationen und ihrer Methodik, sowie zu Evaluationsstandards im Allgemeinen und zu den deutschen Standards der DeGEval im Besonderen zu verzeichnen ist. Auch dass bislang Lernergebnisse und Erfahrungen aus vergangenen Evaluationen kaum für die Optimierung künftiger Projekte genutzt werden, ist kritisch zu sehen.

Die vorliegende Arbeit verfolgte die Zielstellung, diesem Forschungsdefizit entgegenzuwirken, indem eine Meta-Evaluation durchgeführt wurde. Dabei wurde die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ anhand von Evaluationsstandards bewertet. Ein Ziel dieser systematischen Bewertung war es, die Herausforderungen im Evaluationsprozess darzulegen, Empfehlungen dazu zu generieren und so die gemachten Erfahrungen für künftige Studien nutzbar zu machen. Des Weiteren sollten die deutschen Evaluationsstandards unter anderem auf ihre Eignung für Evaluationsbewertungen hin geprüft werden. Und schließlich sollte durch eine detaillierte Beschreibung sowie kritische Reflexion der angewandten Methode ein Beitrag zur fachlichen Diskussion geliefert werden.

Die Evaluationsbewertung wurde durch ein Mitglied des Evaluationsteams und somit in Form einer internen Meta-Evaluation vorgenommen. Grundlage der Bewertung bildeten die deutschen Evaluationsstandards, erweitert um die amerikanischen Richtlinien. Ergänzend zur Bewertung durch die Meta-Evaluatorin wurden Stellungnahmen von Stakeholdern eingeholt, unter anderem, um die Subjektivität zu reduzieren. Aus Gründen der Nachvollziehbarkeit wurde eine ausführliche Schilderung der Bewertungen einschließlich ihrer Begründungen gewählt.

Wie sich durch die Meta-Evaluation zeigte, stellte die Evaluation von „Elternchance ist Kinderchance“ ein Beispiel für eine gelungene Evaluationsstudie dar, anhand welcher vielfältige Herausforderungen im Evaluationsverlauf aufgezeigt werden konnten, die in zahlreichen Empfehlungen für die Evaluationspraxis resultierten. Dabei wurde unter anderem das Spannungsfeld zwischen Wissenschaft und Praxis deutlich, in dem Evaluationen stattfinden. Die DeGEval-Standards

erwiesen sich als sehr nützliches Instrument für die Bewertung von Evaluationen, sie mussten allerdings durch die ausführlicheren Erläuterungen der amerikanischen Evaluationsstandards ergänzt werden. Die Befunde der Meta-Evaluation resultierten in Vorschlägen zur Überarbeitung der DeGEval-Standards. Zur Diskussion um allgemeine vs. feldspezifische Standards sowie um die Möglichkeit der Entwicklung eines umfangreicheren Handbuchs zu den deutschen Richtlinien wurde Stellung genommen. Die angewandte Methode wurde kritisch beleuchtet und unter anderem hinsichtlich der Aspekte der Subjektivität und der Nachvollziehbarkeit diskutiert.

Abstract

Standards for evaluation as well as meta-evaluations (evaluations of evaluation studies) are important measures regarding quality management of evaluations. Therefore, the lack of research addressing meta-evaluations and their methodology, as well as evaluation standards in general and the German Standards for Evaluation by the DeGEval (Evaluation society) in particular, represents a serious issue. Moreover, Lessons Learned and experiences from past evaluations are rarely being used to improve future projects.

The present study was conducted in order to decrease this research gap by implementing a meta-evaluation. The evaluation of the Federal Program "Elternchance ist Kinderchance" ("Opportunities for Parents are Opportunities for Children") was evaluated on the basis of evaluation standards. This systematic assessment aimed at shedding light on the challenges that arose during the evaluation process, generating recommendations and thus making the outcomes usable for future studies. In addition, the German evaluation standards were tested for their applicability in meta-evaluations. Finally, by providing a detailed description and critical reflection of the applied method, a contribution was made to the professional discussion in this area.

The assessment of the evaluation was conducted by a member of the evaluation team and thus as an internal meta-evaluation. It was based on the German Standards for Evaluation, and also on the American Standards for Evaluation. The ratings of the meta-evaluator were complemented by statements of stakeholders. Amongst other benefits, the statements helped reducing the subjectivity of the ratings. In order to reach high comprehensibility, a detailed presentation of the ratings including their rationales was chosen.

As shown by the meta-evaluation, the examined evaluation can be taken as an example of a successful evaluation study, based on which various challenges in the course of the evaluation were identified, which resulted in numerous recommendations for evaluation practice. Amongst other aspects, the tension between science and practice, in which evaluations take place, became apparent. The Standards for Evaluation by the DeGEval proved to be a very useful device for evaluating an evaluation, but they had to be supplemented by the more extensive explanations from the American Standards for Evaluation. The findings of the meta-evaluation resulted in suggestions regarding the

revision of the DeGEval-Standards. The discussion about universal vs. specific evaluation standards and about the possible development of a comprehensive handbook for the German guidelines was commented on. The applied method was critically reflected on and was also discussed, regarding aspects such as subjectivity and comprehensibility.

1 Einleitung

Wissenschaftliche Evaluationen gewinnen in Deutschland mehr und mehr an Bedeutung, wie die Gesellschaft für Evaluation e. V. (DeGEval) feststellt (2008b). Diese Entwicklung ist zu begrüßen, vermögen Evaluationen doch wertvolle Informationen über die evaluierte Maßnahme bereitzustellen und damit beispielsweise für Entscheidungen über die Fortführung oder Beendigung eines Programms eine Grundlage zu liefern. Sie können somit zu einer Kostenreduktion beitragen, da in der Folge weniger Mittel für unwirksame Maßnahmen verschwendet werden (Stockmann, 2002a, 2010).

In Anbetracht der wachsenden Zahl von Evaluationsstudien stellt sich die Frage nach deren Qualität. Denn mangelhaft durchgeführte Evaluationen können im schlimmsten Fall zu fehlerhaften Entscheidungen führen (DeGEval, 2008b). In der Tat finden sich immer wieder Nachweise für Evaluationsstudien von mangelnder Qualität, wie beispielsweise Caspari (2011) für den Bereich der Entwicklungszusammenarbeit und Spiel und Gössler (2001) für das Gebiet der universitären Lehre berichten. Zum Zweck der Qualitätssicherung und -entwicklung wurden Standards für Evaluationen erarbeitet. Diese bieten Evaluatoren[1] eine wichtige Orientierungshilfe für die Planung und Durchführung von Evaluationen. Darüber hinaus können sie bei Evaluationen von Evaluationen (sogenannten *Meta-Evaluationen*) eingesetzt werden, um die Qualität von Evaluationsstudien zu bewerten (DeGEval, 2008b).

Meta-Evaluationen stellen eine wichtige Maßnahme zur Qualitätssicherung von Evaluationen dar (Stufflebeam, 2001b). Allerdings mangelt es bisher an Forschungsarbeiten zu solchen Evaluationsbewertungen und ihrer Methodik. Auch zu den Evaluationsstandards, insbesondere zu den deutschen Standards für Evaluation der DeGEval, wurde bislang wenig Forschung betrieben. So ist beispielsweise unklar, ob diese Standards sich tatsächlich als geeignetes Instrument für Meta-Evaluationen erweisen. Ebenfalls kritisch zu sehen ist die Tatsache, dass bislang Lernergebnisse und Erfahrungen aus vergangenen Evaluationen kaum für die Optimierung künftiger Projekte genutzt werden (Spiel, Popper & Kollmayer, 2012).

[1] In vorliegender Arbeit wird aus Gründen der Lesbarkeit bei gemischtgeschlechtlichen Personengruppen die weibliche Form nicht explizit angeführt.

Die vorliegende Arbeit hat sich zum Ziel gesetzt, einen Beitrag zur Schließung der beschriebenen Forschungslücken zu leisten. Sie stellt eine Meta-Evaluation vor, bei der die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ anhand von Evaluationsstandards bewertet wurde. Dabei steht nicht das Bewertungsergebnis im Fokus, vielmehr werden anhand der systematischen Bewertung die Herausforderungen dieser Evaluation erarbeitet und die gemachten Erfahrungen für künftige Studien nutzbar gemacht. Die Evaluationsstandards der DeGEval werden auf ihre Eignung für Evaluationsbewertungen geprüft und auch im Hinblick auf weitere Aspekte bewertet. Darüber hinaus wird die angewandte Methode der Meta-Evaluation möglichst umfassend beschrieben und zudem kritisch reflektiert, um die Diskussion in diesem Bereich zu fördern und zur Qualitätsverbesserung von Evaluationsbewertungen beizutragen.

Die Arbeit ist aufgebaut wie folgt: Zunächst wird im theoretischen Teil eine Einführung in den Themenbereich der Evaluationsforschung gegeben. Nach einer Begriffsbestimmung wird auf Merkmale von Evaluationsstudien und auf die Phasen eines Evaluationsprozesses eingegangen, Bedeutung und Grenzen von Evaluationen werden erläutert und ihre Nutzung wird erörtert. Zuletzt wird das Spannungsfeld zwischen Wissenschaft und Praxis thematisiert, in dem Evaluationen sich befinden. Im zweiten Abschnitt des Theorieteils werden Evaluationsstandards und Meta-Evaluationen behandelt. Nach einer Einführung in das Feld der Evaluationsstandards werden die Richtlinien der DeGEval genauer vorgestellt. Auf einige grundlegende Informationen zu Meta-Evaluationen folgt schließlich ein Überblick zum aktuellen Forschungsstand hinsichtlich Meta-Evaluationen und Evaluationsstandards, welcher auf die Ziele und Fragestellungen der vorliegenden Arbeit hinleitet.

Anschließend wird die Meta-Evaluation dargestellt. Dazu erfolgt zuerst eine Schilderung des Bundesprogramms „Elternchance ist Kinderchance“ und seiner Evaluation. Dann wird die Methode der Meta-Evaluation beschrieben und danach die Evaluationsbewertung anhand von 25 Evaluationsstandards vorgenommen. Dabei werden neben den Anforderungen des jeweiligen Standards auch die zur Bewertung herangezogenen Daten möglichst detailliert angeführt, um die Beurteilung nachvollziehbar zu machen.

Durchführung und Ergebnisse der Meta-Evaluation werden im Diskussionsteil hinsichtlich dreier Gesichtspunkte analysiert. Zunächst werden die zentralen Herausforderungen im Evaluationsprozess herausgearbeitet und daraus Empfehlungen für Evaluatoren und Auftraggeber von Evaluationen abgeleitet. Es folgen eine Bewertung der DeGEval-Standards sowie Überlegungen zu ihrer möglichen Überarbeitung im Rahmen des derzeit laufenden Revisionsprozesses. Danach wird die Methode der Meta-Evaluation kritisch reflektiert. Abschließend wird ein Fazit zur Studie gezogen und auf Forschungsimplikationen hingewiesen.

2 Theoretischer Hintergrund

2.1 Evaluation

Zunächst wird eine Einführung in das Themengebiet der Evaluation gegeben. Diese beginnt mit einer Begriffsbestimmung und einem kurzen Abriss der Entwicklung der Evaluationsforschung. Es werden einige Merkmale zur Einordnung von Evaluationsstudien vorgestellt und der Ablauf einer Evaluation wird anhand eines Phasenmodells veranschaulicht. Weiterhin werden Bedeutung und Grenzen von Evaluationen aufgezeigt und es wird auf ihren Nutzen sowie auf ihre Nutzung eingegangen. Abschließend wird das Spannungsfeld zwischen Wissenschaft und Praxis erörtert, in dem Evaluationen stattfinden.

2.1.1 Definition und Begriffsklärung

Zum Terminus der Evaluation finden sich in der Literatur einerseits vielfältige Begriffsbestimmungen, andererseits bleibt er dennoch teilweise unscharf (vgl. z. B. Hornbostel, 2010; Wottawa & Thierau, 1998). Einführend erfolgt daher zunächst eine Definition von Evaluation, zudem wird geklärt, welcher Evaluationsbegriff der vorliegenden Arbeit zugrunde liegt.

Der Ausdruck *Evaluation* stammt aus dem Lateinischen und enthält das Wort „valor“ (Wert) und das Präfix „e“ bzw. „ex“ (aus). Wörtlich ergibt sich daraus die Bedeutung „einen Wert aus etwas ziehen“, also zunächst einmal eine Bewertung jedweder Art (Stockmann, 2007). Eine solch breit gefasste Definition umfasst somit auch jegliche alltägliche Ermittlung des Wertes von Dingen oder Sachverhalten.

Von diesem weiten Evaluationsverständnis abzugrenzen ist der Begriff der Evaluationsforschung bzw. der wissenschaftlichen Evaluation, bei welcher die Festlegung eines Wertes auf systematischer Datenerhebung und -analyse beruht. Einer anerkannten Definition zufolge ist Evaluationsforschung die „systematische Anwendung sozialwissenschaftlicher Forschungsmethoden zur Beurteilung der Konzeption, Ausgestaltung, Umsetzung und des Nutzens sozialer Interventionsprogramme“ (Rossi, Freeman & Hofmann, 1988, S. 3). Zu beachten ist, dass die Begriffe *Evaluation* und *Evaluationsforschung* in der Forschungsliteratur weitgehend synonym verwendet werden (Atria, Reimann &

Spiel, 2006; Wottawa & Thierau, 1998). Auch in der vorliegenden Studie ist mit dem Begriff „Evaluation“ stets der wissenschaftliche, datengestützte Wortsinn gemeint.

Evaluationsstudien sind prinzipiell der angewandten Sozialforschung zuzuordnen (Stockmann, 2007) und unterscheiden sich von der Grundlagenforschung in mehreren Aspekten, wie in Abschnitt 2.1.7.1 näher erläutert wird. Zu erwähnen ist zudem die Komplexität von Evaluationen, welche sich unter anderem darauf gründet, dass diese in verschiedensten Praxisfeldern durchgeführt werden und vielfältige Gegenstände behandeln, sowie unterschiedliche Funktionen haben können (Böttcher, Kerlen, Maats, Schwab & Sheikh, 2014a). Einige Kriterien zur Einordnung von Evaluationen werden in Abschnitt 2.1.3 behandelt. Zunächst soll jedoch ein kurzer Überblick über die Entwicklung der Evaluationsforschung gegeben werden.

2.1.2 Entwicklung der Evaluationsforschung

Die wissenschaftliche Evaluation hat, wie Moosbrugger und Schweizer es treffend ausdrücken, „eine lange Vergangenheit, doch nur eine kurze Geschichte“ (2002, S. 19). So werden die Ursprünge der Evaluationsforschung teils im 18. Jahrhundert, teils sogar bereits in der Steinzeit festgemacht, und auch in der Bibel finden sich Beispiele (Fitzpatrick, Sanders & Worthen, 2012; Meyer, 2002; Rossi et al., 1988; Wottawa & Thierau, 1998).

Als Ausgangspunkt der modernen (Programm-)Evaluation in den USA gelten im Allgemeinen die 1960er-Jahre. Zuvor waren z. B. in Schulen und im Gesundheitssektor Evaluationsstudien durchgeführt worden, im zweiten Weltkrieg bestand ein Tätigkeitsfeld darin, die Wirksamkeit verschiedener Propaganda-Methoden zu prüfen. Doch erst die umfangreichen Reformen in den 1960er-Jahren brachten einen enormen Entwicklungsschub für die wissenschaftliche Evaluation mit sich, da die Reformprogramme zum Zwecke der Kontrolle und Verbesserung, aber auch der politischen Rechtfertigung evaluiert wurden. Die positive Entwicklung im Evaluationsbereich ließ sich zudem auf Fortschritte in den empirischen Forschungsmethoden zurückführen, darunter auch die Verbesserung statistischer Auswertungsverfahren. Im Zuge der nun florierenden Evaluationsforschung wurden ab den 1970er-Jahren Lehrbücher über Evaluation publiziert, Fachzeitschriften

gegründet sowie Evaluationsgesellschaften ins Leben gerufen. Mit universitären Studiengängen, Berufsverbänden und Institutionen hat sich in den USA heute der eigenständige Berufsstand der Evaluatoren etabliert (Fitzpatrick et al., 2012; Kury & Obergfell-Fuchs, 2012; Rossi et al., 1988; Stockmann, 2002b; Wottawa & Thierau, 1998).

In Deutschland hingegen begann die Entwicklung der wissenschaftlichen Evaluation später (etwa ab den 1970er-Jahren) und erlebte auch keinen so großen Aufschwung, wie dies in den USA der Fall war. Die Evaluationsforschung entwickelte sich zwar langsamer, aber dennoch kontinuierlich und gewann in verschiedenen Bereichen zunehmend an Bedeutung (Kury & Obergfell-Fuchs, 2012; Rossi et al., 1988). Diese Fortschritte zeigten sich auch darin, dass im Jahr 1997 die DeGEval sowie im Jahr 2002 die deutschsprachige Zeitschrift für Evaluation gegründet wurde (Stockmann, 2002a). Heute werden Evaluationen stark nachgefragt, da es in Zeiten knapper Ressourcen die Wirksamkeit und Nachhaltigkeit (politischer) Programme zu überprüfen gilt. Jedoch geschieht Evaluationsforschung meist sektoral, interdisziplinärer Austausch findet kaum statt. Auch die Professionalisierung ist im Vergleich mit den USA noch nicht weit gediehen. So gibt es so gut wie keine eigenständigen Studiengänge zum Thema Evaluation. Einzelne Veranstaltungen werden z. B. im Rahmen eines erziehungs-, sozial- oder wirtschaftswissenschaftlichen Studiums angeboten, und lediglich in der Psychologie ist die Evaluationsforschung im Modul „Forschungsmethoden“ fester Bestandteil der Curricula (Hense & Steckhan, 2014; Moosbrugger & Schweizer, 2002; Stockmann, 2002b). Die Thematik der Professionalisierung von Evaluation wird derzeit diskutiert (vgl. Abschnitt 2.1.7.3).

Im Verlauf der Entwicklung, die die Evaluationsforschung durchlaufen hat, hat sich in methodischer Hinsicht ein Wandel von rein quantitativer zu qualitativer Forschung bzw. zu einer Kombination im Sinne eines *Mixed-Methods*-Ansatzes vollzogen. Eine weitere Veränderung besteht darin, dass Evaluatoren nicht mehr lediglich über methodische Fertigkeiten verfügen müssen, sondern zu ihrem Tätigkeitsfeld auch Management, soziale Kompetenzen sowie teilweise politische Handlungen gerechnet werden (vgl. Rossi et al., 1988; Ulrich & Wenzel, 2003). Evaluierende führen somit keine reine Forschungstätigkeit durch, sondern befinden sich mitunter in einem Spannungsfeld zwischen Wissenschaft und Praxis bzw. Politik, was in Abschnitt 2.1.7 erörtert wird.

2.1.3 Charakteristika zur Einordnung von Evaluationen

In der Forschungsliteratur finden sich viele Merkmale, anhand derer sich Evaluationen beschreiben und einordnen lassen. Im Folgenden werden einige dieser Charakteristika behandelt, wobei die Darstellung nicht auf Vollständigkeit abzielt, sondern die Relevanz in Bezug auf die später vorgestellte eigene Evaluationsstudie bedeutsam ist. In den folgenden Abschnitten werden zunächst Anwendungsbereiche und Gegenstände von Evaluationen behandelt, danach werden Evaluationsziele und -funktionen dargestellt. Es folgt eine Einteilung nach dem Zeitpunkt der Evaluation. Danach wird auf die Unterscheidung zwischen interner und externer Evaluation eingegangen und abschließend werden verschiedene Evaluationsansätze erörtert.

2.1.3.1 Anwendungsbereiche und Gegenstände von Evaluationen

Evaluationsstudien werden in diversen Praxisfeldern durchgeführt, z. B. im Bildungssektor, im Gesundheitswesen oder auch in der Arbeitsmarktpolitik (Wottawa & Thierau, 1998). Innerhalb der Psychologie ist die Evaluationsforschung laut Moosbrugger und Schweizer (2002) in folgenden Bereichen von besonderer Bedeutung: Evaluation sozialer Programme, Therapieevaluation, Evaluation von Maßnahmen im Kontext der Arbeits- und Organisationspsychologie, Evaluation im pädagogischen Bereich sowie Lehrevaluation. Die Vielfältigkeit der Arbeitsfelder von Evaluatoren mit ihren jeweils unterschiedlichen Anforderungen zeigt sich auch darin, dass sich von den 16 Arbeitskreisen der DeGEval der überwiegende Anteil spezifischen Praxisfeldern widmet (z. B. Entwicklungspolitik, Kultur, Schule) und vergleichsweise wenige Themen in querschnittlichen Arbeitskreisen behandelt werden (z. B. Methoden; Böttcher, Kerlen, Maats, Schwab & Sheikh, 2014b). Für einen aktuellen Überblick über die verschiedenen Evaluationsfelder in Deutschland und Österreich sowie den derzeitigen Stand und Entwicklungsperspektiven sei an dieser Stelle auf die DeGEval-Publikation von Böttcher und Kollegen verwiesen (Böttcher et al., 2014b).

Neben den Anwendungsbereichen lassen sich auch verschiedene Evaluationsgegenstände bzw. -objekte differenzieren. Wottawa und Thierau (1998) zählen hierzu Personen, Umweltfaktoren, Produkte, Methoden, Zielvorgaben, Programme, Projekte, Systeme und Forschungsergebnisse. Der letztgenannte Evaluationsgegenstand umfasst

auch Evaluationsstudien, deren Evaluierung als *Meta-Evaluation* bezeichnet wird. Darauf wird in Abschnitt 2.2 näher eingegangen. Von der Art des Evaluationsobjektes sowie auch vom Anwendungsbereich hängt ab, wie eine Evaluation angemessen zu konzipieren und durchzuführen ist (Widmer, 2008).

2.1.3.2 Ziele und Funktionen von Evaluationen

Evaluationen können verschiedene Zielsetzungen verfolgen und daher hinsichtlich ihrer Funktion eingeteilt werden. Die in der einschlägigen Literatur berichteten Taxonomien umfassen für gewöhnlich zwischen drei und fünf Evaluationsfunktionen oder auch -zwecke. Da diese sich inhaltlich nicht deutlich voneinander unterscheiden, wird an dieser Stelle lediglich die detaillierteste Einteilung vorgestellt. Bortz und Döring (2006) beschreiben die folgenden fünf zentralen Funktionen von Evaluationen (vgl. auch Stockmann, 2007; Wottawa & Thierau, 1998):

1. **Erkenntnisfunktion**: Evaluationsstudien stellen wissenschaftliche Erkenntnisse zum Evaluationsgegenstand, dem Evaluandum, bereit und tragen somit zur Sammlung von Wissen bei.
2. **Optimierungsfunktion**: Hier wird die Zielstellung verfolgt, das Evaluandum zu optimieren, also die Stärken z. B. eines Programms besser zu nutzen und seine Schwächen auszugleichen.
3. **Kontrollfunktion**: Evaluationen können aufzeigen, ob eine Maßnahme korrekt implementiert wird und inwieweit die erhofften Wirkungen erreicht werden. Zudem können sie unbeabsichtigte (positive wie negative) Nebenwirkungen offenbaren.
4. **Entscheidungsfunktion**: Ein weiteres Ziel von Evaluationen besteht darin, Informationen zu liefern, auf deren Grundlage Entscheidungen getroffen werden, z. B. darüber, ob ein Programm fortgeführt oder eingestellt, oder ob es weiterentwickelt wird. Eine andere mögliche Entscheidung kann die Wahl zwischen zwei konkurrierenden Programmen darstellen, welche zu diesem Zweck evaluiert werden.

5. **Legitimationsfunktion**: Schließlich können Evaluationen selbst und ihre Befunde dazu dienen, die evaluierte Maßnahme zu legitimieren und verwendete Mittel zu rechtfertigen.

Evaluationsstudien können in der Praxis eine dieser Funktionen erfüllen, es können aber auch mehrere Zwecke gleichzeitig oder nacheinander im Verlauf eines Projektes verfolgt werden. Tatsächlich wird die Erkenntnisfunktion üblicherweise auch dann erfüllt, wenn ein anderes Ziel im Vordergrund steht, da Evaluationsstudien immer einen Erkenntnisgewinn mit sich bringen (Holling, 2009; Stockmann, 2007). Wichtig ist, dass die Evaluationszwecke klar definiert sind, da sich nur dann eindeutige Ziele ableiten sowie Operationalisierungen erarbeiten lassen und das Evaluationsteam den Evaluationsprozess auf den jeweiligen Zweck hin ausrichten kann. In der Realität kommt es jedoch häufig vor, dass die Funktion der Evaluation nicht explizit gemacht wird und die Ziele nicht konkret benannt werden können, was bei der Realisierung von Evaluationsprojekten störend und hinderlich sein kann (Beywl, Kehr, Mäder & Niestroj, 2007; Bortz & Döring, 2006; Stockmann, 2011; Wottawa & Thierau, 1998).

Die hier dargestellten fünf Evaluationsfunktionen sind solche, die (zumindest idealerweise) auf Ergebnisoffenheit abzielen und sich daher mit der Berufsethik von Wissenschaftlern am ehesten vereinbaren lassen. Neben diesen „Lehrbuchfunktionen“ kommt es im Evaluationsalltag jedoch mitunter auch dazu, dass Evaluationen instrumentalisiert werden und dann strategische Funktionen erfüllen. So werden Evaluationen manchmal durchgeführt, um die Motivation und Leistungen der evaluierten Personen zu steigern, wenn diese wissen, dass sie überprüft werden. Auch werden die erhobenen Daten teilweise nicht näher ausgewertet, geschweige denn für die Praxis genutzt, wenn eine Evaluation lediglich nach außen Kunden- oder Mitarbeiterorientierung
demonstrieren soll. Ein weiterer strategischer Zweck besteht darin, Evaluationen als Durchsetzungshilfe zu verwenden. Besonders bei Entscheidungen, die vermutlich auf wenig Akzeptanz stoßen werden, kann die Verantwortung an die Evaluatoren delegiert werden. Dies birgt natürlich die Gefahr, dass die Ergebnisgewinnung von den Entscheidungsträgern in die gewünschte Richtung beeinflusst wird. Mitunter können Evaluationen sogar dazu dienen, im Nachhinein Entscheidungen zu rechtfertigen, die bereits getroffen wurden. Und

schließlich kann es vorkommen, dass Evaluationen rein symbolischen Wert haben, da sie nur beauftragt werden, weil es entsprechende Vorschriften zu erfüllen gilt (vgl. Bortz & Döring, 2006; Holling, 2009; Spiel, Gradinger & Lüftenegger, 2010; Stockmann, 2007). Solche Instrumentalisierungen von Evaluationen sind zwar vermutlich eher die Ausnahme, in der Regel verfolgen Auftraggebende sicherlich sinnvolle und moralisch vertretbare Zwecksetzungen. An Evaluationen beteiligte Personen sollten sich jedoch dessen bewusst sein, dass es auch missbräuchliche Verwendungen gibt, zumal diese meist nicht offengelegt werden.

Neben der eben skizzierten Klassifikation von Evaluationszwecken besteht eine weitere gängige Einteilung in der Unterscheidung zwischen formativer und summativer Evaluation (nach Michael Scriven, 1967, 1972, 1991; zitiert nach Mittag & Hager, 2000). Formative (gestaltende) Evaluation zielt auf die Optimierung des Evaluandums ab. Sie wird daher in der Regel während der Entwicklung und Erprobung einer Maßnahme durchgeführt und beruht gewöhnlich auf mehrfachen Datenerhebungen im Verlauf. Ihre Aufgabe besteht darin, Stärken und Schwächen aufzuzeigen, damit Schwächen künftig vermieden und Stärken besser genutzt werden können. Summative (zusammenfassende) Evaluationen haben den Zweck, die Wirksamkeit des Evaluandums zu untersuchen, also z. B. die Frage, inwieweit erwartete Wirkungen eines Programms eintreten oder wie kosteneffizient es ist. Die Ergebnisse können dazu dienen, die Anwendung einer Maßnahme zu begründen, sie können aber auch als Entscheidungsgrundlage z. B. dafür genutzt werden, ob ein Programm weitergeführt, ausgeweitet oder abgebrochen werden soll. Im besten Fall umfasst eine Evaluation sowohl den formativen als auch den summativen Ansatz, damit beispielsweise Präventionsprogramme ihre Ziele bestmöglich erreichen können (Fitzpatrick et al., 2012; Holling, 2009; Mittag, 2003; Mittag & Hager, 2000; Znoj & Regli, 2006).

In Bezug auf die zuvor dargestellten fünf Funktionen von Evaluationen sei abschließend erwähnt, dass deutliche Überschneidungspunkte beider Klassifikationsmöglichkeiten erkennbar sind. So findet sich die formative Zielsetzung in der Optimierungsfunktion wieder, während summative Evaluationen oftmals die Grundlage für Entscheidungen liefern sollen und somit Parallelen zur Entscheidungs-, aber auch zur Legitimations- und zur Kontrollfunktion gezogen werden

können. Und sowohl formative als auch summative Evaluationen tragen selbstverständlich zum wissenschaftlichen Erkenntnisgewinn bei.

2.1.3.3 Zeitpunkt der Evaluation

Wissenschaftliche Evaluationen können auch danach eingeteilt werden, zu welchem Zeitpunkt im Verlauf einer Maßnahme sie durchgeführt werden. Mittag und Hager (2000) erörtern in Bezug auf psychologische Interventionsprogramme die folgenden Evaluationsarten (vgl. auch Moosbrugger & Schweizer, 2002; Rossi et al., 1988):

1. **Evaluation der Programmkonzeption**: Diese findet zum frühestmöglichen Zeitpunkt statt, während der Entwicklung eines Programmes, und wird daher auch antizipatorische oder prospektive Evaluation genannt. Es soll geklärt werden, welchen Bereich das Programm abdecken und welche Ziele es verfolgen soll. Zudem ist die Konzeption daraufhin zu überprüfen, ob sie zur Zielerreichung geeignet ist.
2. **Formative Evaluation**: In der sich anschließenden Erprobungsphase des Programms ist die Aufgabe der formativen Evaluation, Verbesserungsmöglichkeiten hinsichtlich Zielsetzung, Anwendbarkeit und Wirksamkeit zu erarbeiten und die Maßnahme entsprechend zu optimieren. Werden dabei wesentliche Programmkomponenten verändert, kann dies eine erneute Überprüfung notwendig machen.
3. **Evaluation der Programmdurchführung**: Nun wird das Programm im Alltag angewendet und es wird erforscht, wie diese Implementation gelingt. Dabei ist von Interesse, inwiefern die Maßnahme unter Alltagsbedingungen realisiert werden kann und wo möglicherweise Probleme auftreten. Auch die Teilnehmenden am Programm werden untersucht, wobei besonders auf Verzerrungen der Programmreichweite (inwieweit wird die Zielgruppe tatsächlich erreicht?) und Abbrecher zu achten ist.
4. **Evaluation der Programmwirksamkeit**: Hier ist zu unterscheiden zwischen der Prozessevaluation, bei welcher Veränderungen im Verlauf des Programmes analysiert werden, und der Erfolgs- bzw. Ergebnisevaluation, welche sich mit Veränderungen nach der Programmdurchführung beschäftigt.

Neben der Frage, ob ein Programm wirksam ist, ist auch wichtig, zu überprüfen, wie intensiv die Wirkungen sind und welche Nebenwirkungen existieren.

5. **Evaluation der Programmeffizienz**: Ebenfalls nach der Durchführung des Programms angesiedelt, erfolgt die Bewertung der Programmeffizienz zuletzt. Denn um die Wirtschaftlichkeit eines Programmes einschätzen zu können, müssen zuvor die Wirkungen untersucht werden, um diese gegen die entstandenen Kosten abzuwägen.

Die Evaluationsarten im zeitlichen Verlauf eines Interventionsprogramms veranschaulicht Abbildung 1.

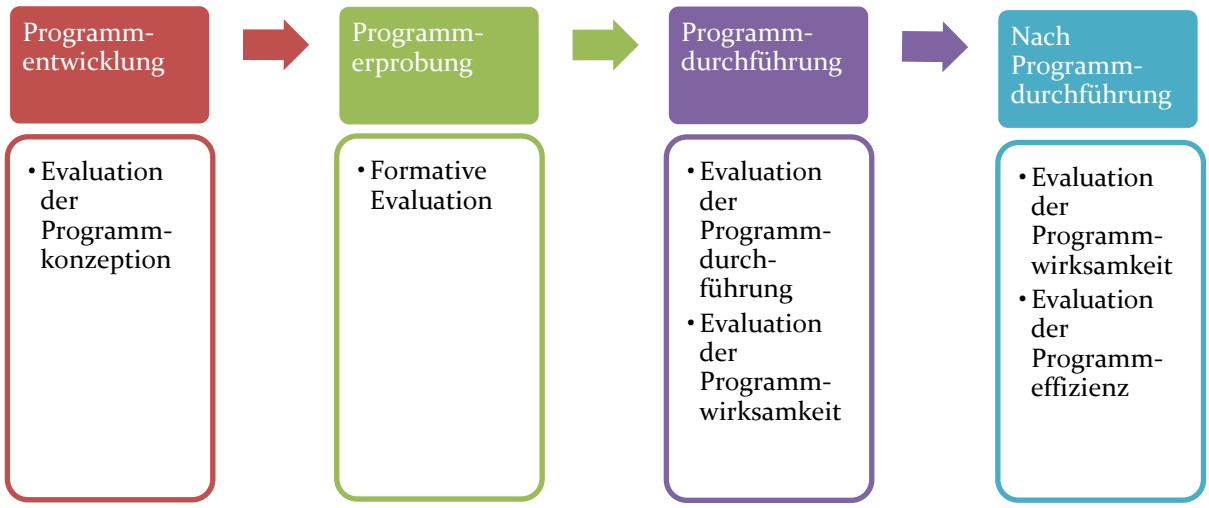


Abbildung 1: Evaluationsarten zu verschiedenen Zeitpunkten im Programmablauf (nach Mittag & Hager, 2000, S. 108)

2.1.3.4 Interne und externe Evaluationen

Als ein weiteres Merkmal von Evaluationen sei die Rolle der Evaluierenden genannt. So können Evaluationen intern durchgeführt werden, d. h. von Mitarbeitern der Organisation, die auch die zu evaluierende Maßnahme anbietet. Sind die Evaluatoren dabei in der gleichen Abteilung tätig, welche auch für die Durchführung der Maßnahme zuständig ist, dann wird diese Form der Evaluation als *Selbstevaluation* bezeichnet. Es besteht demgegenüber auch die Möglichkeit, externe

Evaluatoren hinzuzuziehen, welche nicht dem Anbieter der Maßnahme angehören (Fitzpatrick et al., 2012; Rossi et al., 1988; Stockmann, 2007).

Beide Evaluationsformen haben Vor- und Nachteile. Bei internen Evaluationen ist von Vorteil, dass die Evaluierenden durch ihre Einbindung in die Organisation vertiefte inhaltliche Einblicke haben und damit eine hohe Sachkenntnis aufweisen. Zudem ist die Durchführung interner Evaluationen mit weniger Aufwand verbunden und daher insgesamt schneller realisierbar, auch die Ergebnisse können direkt in die Praxis umgesetzt werden. Allerdings besitzen interne Evaluatoren gewöhnlich keine ausreichenden Methodenkenntnisse, wie sie für eine wissenschaftliche Evaluation erforderlich sind. Ein weiterer Nachteil liegt in der Verbundenheit zum Evaluandum bzw. zur durchführenden Organisation begründet, da die offensichtliche Gefahr besteht, dass die Evaluierenden nicht genug Distanz aufweisen, um die Maßnahme unvoreingenommen bewerten zu können. Auch geeignete Alternativen zum evaluierten Programm könnten dadurch nicht ausreichend berücksichtigt werden (ebenda).

Die genannten Vor- und Nachteile interner Evaluationen finden sich gespiegelt bei externen Evaluationen wieder. So zeichnen sich externe Evaluatoren durch größere Unabhängigkeit aus, welche ihnen wiederum zu hoher Glaubwürdigkeit verhilft. Zudem verfügen sie in der Regel über eine fundierte methodische Ausbildung. Während sie zwar den Fachbereich des Evaluandums kennen oder sich gezielt einarbeiten, besitzen sie allerdings keine so detaillierte Kenntnis der Maßnahme und/oder Organisation wie dies bei internen Evaluierenden der Fall ist. Ein weiterer Nachteil bei externen Evaluationen besteht darin, dass die Evaluierenden von außen kommen und bei den Programmdurchführenden oder anderweitig von der Evaluation betroffenen Personen nicht grundsätzlich auf Akzeptanz stoßen, sondern vielmehr sogar Ängste und Abwehrreaktionen auslösen können. Dies kann die Durchführung der Evaluation sowie die Umsetzung der Ergebnisse erschweren (ebenda).

Bezüglich der Kosten, die die Beauftragung externer Evaluatoren mit sich bringt, ist zu bemerken, dass diese nicht grundsätzlich teurer sein muss als die Durchführung einer internen Evaluation, da auch hierfür Kosten anfallen. Stockmann (2007) kommt zu dem Schluss, dass die finanziellen Aufwendungen bei externen und internen Evaluationen ähnlich hoch ausfallen und gerade für kleinere Unternehmen

externe Evaluationen aus ökonomischen Gesichtspunkten sogar deutlich rentabler sein können.

Hinsichtlich der Unabhängigkeit externer Evaluatoren ist zu beachten, dass auch diese in einem Abhängigkeitsverhältnis stehen, da sie von einem Geldgeber finanziert werden, der für gewöhnlich gewisse Interessen bezüglich der evaluierten Maßnahme verfolgt. Hier kann es durchaus zu einem Spannungsfeld kommen zwischen dem Anspruch, unabhängige Forschung zu betreiben und dem Bewusstsein, dass von den Ergebnissen dieser Evaluation möglicherweise abhängt, ob der Geldgeber den Evaluatoren weiterhin lukrative Aufträge erteilt (Brandt, 2007). Abschnitt 2.1.7.1 beleuchtet die Herausforderungen der Auftragsforschung näher.

Zusammenfassend ist festzuhalten, dass interne und externe Evaluationen jeweils Vor- und Nachteile mit sich bringen. In der Praxis bietet es sich daher an, beide Ansätze zu kombinieren, um die Nachteile ausgleichen und von den Vorteilen besonders profitieren zu können (Fitzpatrick et al., 2012; Rossi et al., 1988; Stockmann, 2007).

2.1.3.5 Evaluationsansätze

Schließlich können verschiedene Evaluationsansätze bzw. Evaluationsmodelle unterschieden werden. Evaluationsansätze existieren mittlerweile in einer solchen Vielzahl, dass ihre Kategorisierung sich sehr schwierig gestaltet. Als Grund für die Komplexität bestehender Ansätze und Modelle nennt Stockmann (2007) unter anderem die Tatsache, dass bei vielen Evaluationsprojekten keine Zuordnung zu vorhandenen Modellen vorgenommen wird. Dies führt dazu, dass oftmals eigene Vorgehensweisen für Evaluationen entwickelt werden, die allerdings durchaus Überschneidungen mit bereits bestehenden Ansätzen erkennen lassen. Im Folgenden wird eine Typologie vorgestellt, anhand derer Evaluationsansätze fünf verschiedenen Gruppen zugeordnet werden können (vgl. Fitzpatrick et al., 2012; Stockmann, 2007):

- **Zielorientierte Ansätze**: Hierbei steht die Frage nach der Zielerreichung von Programmen und Maßnahmen im Mittelpunkt. Aus den Ergebnissen können Optimierungsmöglichkeiten für die Programmprozesse oder Anpassungen der Ziele abgeleitet werden.

- **Managementorientierte Ansätze**: Zentral sind hier die Informationsbedürfnisse der Entscheidungsträger. Diese sind zu identifizieren und zu erfüllen, indem den Entscheidern zeitgerecht Informationen z. B. über Programmabläufe zur Verfügung gestellt werden.
- **Konsumentenorientierte Ansätze**: Bei diesen Ansätzen liegt der Fokus auf Verbrauchern, die mit produktbezogenen Informationen und Bewertungen versorgt werden sollen. Dabei kommen häufig Checklisten zum Einsatz.
- **Expertenorientierte Ansätze**: Bei Evaluationen, die diesem Ansatz folgen, wird die Bewertung durch Experten vorgenommen.
- **Partizipative Ansätze**: Hier stehen die an einer Evaluation beteiligten oder von ihr betroffenen Personengruppen, die sogenannten Stakeholder, im Zentrum. Sie sollen bei der Planung und Durchführung der Evaluation einbezogen werden.

Zu beachten ist, dass mitunter keine eindeutige bzw. exklusive Einordnung vorgenommen werden kann, da Evaluationen Merkmale mehrerer Kategorien aufweisen können. Für ausführlichere Erläuterungen zu den fünf Kategorien sowie für alternative Einteilungen von Evaluationsansätzen sei der Leser auf die einschlägige Literatur verwiesen (z. B. Beywl & Speer, 2004; Fitzpatrick et al., 2012; Stockmann, 2007).

2.1.4 Phasen eines Evaluationsprozesses

Um den Evaluationsprozess zu veranschaulichen, wird in diesem Abschnitt ein Überblick über den Ablauf einer Evaluation und die in jeder Phase üblicherweise anfallenden Aufgaben gegeben. Die Darstellung erfolgt in Anlehnung an Dörings Phasenmodell (Döring, 2009) sowie die von Beywl und Kollegen (Beywl et al., 2007) beschriebenen Evaluationsschritte (vgl. auch Rost, 2000; Wottawa & Thierau, 1998). In Abbildung 2 sind die Evaluationsphasen zusammengefasst, die im Folgenden näher erläutert werden.

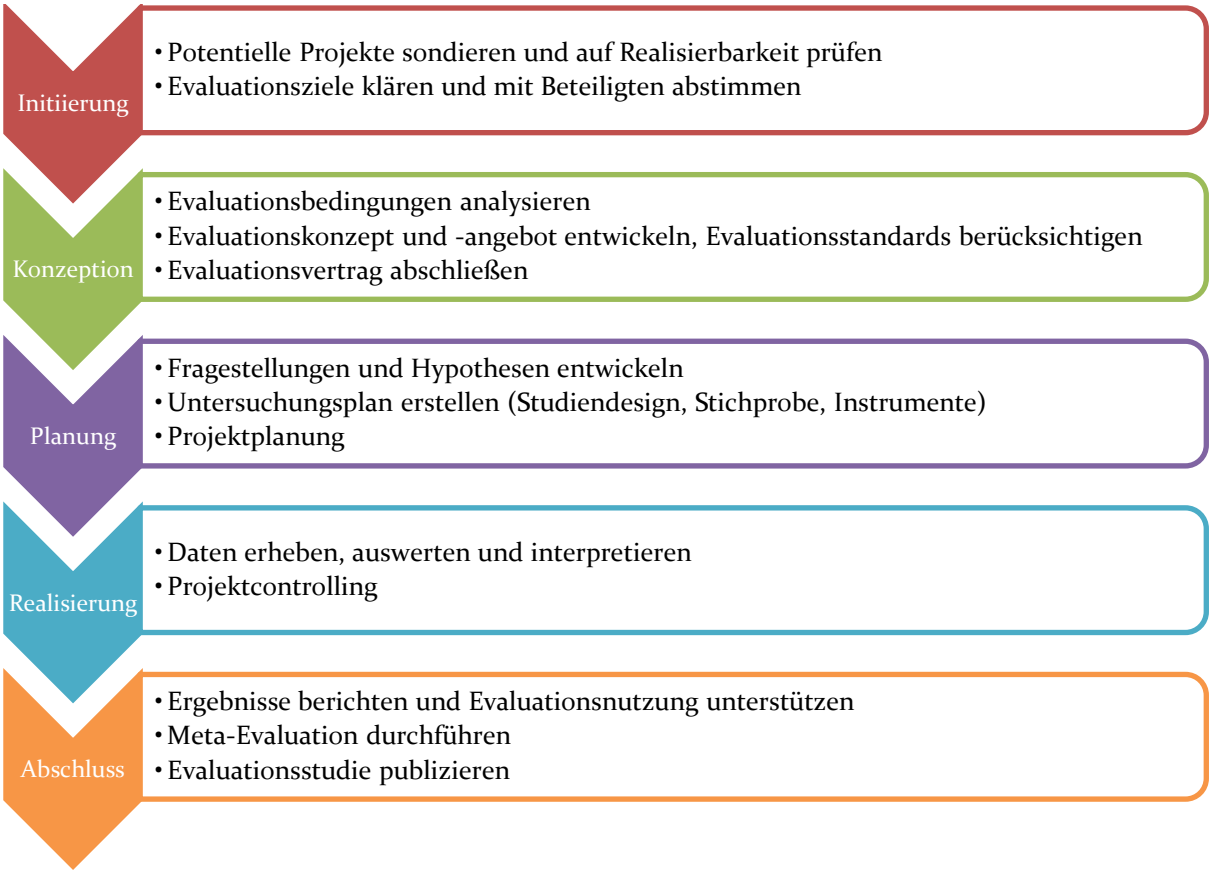


Abbildung 2: Evaluationsphasen und -aufgaben
(nach Döring, 2009, S. 104)

Initiierungsphase. Evaluationen beginnen für gewöhnlich mit einer Initiierungsphase, in welcher Evaluierende von einem potentiellen Projekt erfahren, dieses prüfen und sich gegebenenfalls dafür bewerben. Evaluationen werden z. B. über Ausschreibungen vergeben, woraufhin Evaluatoren Angebote einreichen und somit im Wettbewerb zueinander stehen. Teilweise werden auch gezielt bestimmte Evaluatoren angefragt, mit denen der Auftraggeber gute Erfahrungen gemacht hat oder die ein gutes Renommee genießen. Evaluatoren sollten jedes Projekt darauf prüfen, ob es zeitlich realistisch umsetzbar ist und zu den eigenen Kompetenzen passt. In einem weiteren Schritt gilt es, die Evaluationsfunktion zu klären, um Ziele und Operationalisierungen ableiten zu können. Wie bereits erwähnt, kann dies in der Praxis schwierig sein, da viele Auftraggebenden keine konkreten Vorstellungen von den Zielen der Evaluation oder gar des Evaluandums haben (vgl. Abschnitt 2.1.3.2). Evaluatoren müssen häufig vermitteln, dass Kausalitäten schwer nachzuweisen sind und Auftraggebende ihre Erwartungen an die Forschungsrealität anpassen sollten.

Konzeptionsphase. Haben Evaluatoren ein Projekt ermittelt, das ihnen realisierbar erscheint, folgt die Konzeptionsphase, in der zunächst die Evaluationsbedingungen analysiert werden. Dabei stellt die Zusammensetzung des Evaluationsteams einen zentralen Aspekt dar. Der Umfang kann zwischen einer einzigen Person und mehr als einem Dutzend variieren und ist abhängig vom Umfang der Evaluation und von den erforderlichen Kompetenzen (vgl. Abschnitt 2.1.7.3). Wichtig ist zudem, dass die Evaluatoren sich ein klares Bild vom Evaluationsgegenstand machen, also die Merkmale des Evaluandums und seiner Nutzer kennen und die Erwartungen des Auftraggebers klären.

Nur mit genauem Hintergrundwissen ausgestattet ist es möglich, ein passendes Evaluationskonzept zu erarbeiten und darauf basierend ein Angebot vorzulegen, um den Zuschlag für das Projekt erhalten zu können. Das Evaluationskonzept soll unter anderem die Evaluationsziele, Forschungsfragen und Hypothesen, aber auch Angaben zum Studiendesign, zur Datenanalyse und zur Berichterstattung enthalten. Auch das Projektmanagement, welches Zeit- und Personalplanung, Projektbudget sowie organisatorische und Kommunikationsaufgaben umfasst, soll hier abgebildet werden.

Zu beachten ist, dass das Evaluationskonzept noch nicht in allen Punkten vollständig sein kann, da einerseits das Angebot zu einem Zeitpunkt entwickelt wird, zu dem noch nicht klar ist, ob die Anbietenden den Zuschlag erhalten werden. Da sie somit unbezahlt in Vorleistung gehen, ist aus ökonomischen Gründen eine gewisse Beschränkung des Arbeitsaufwands geboten und nachvollziehbar. Einige Angaben, z. B. zum Vorgehen bei der Datenerhebung, werden daher im Detail erst nach Vertragsabschluss vereinbart.

Andererseits verfügen die Evaluatoren in der Regel noch nicht über alle notwendigen Informationen und Kontakte und können daher nicht jeden Arbeitsschritt mit Sicherheit planen. Beispielsweise gehört es zu ihren Aufgaben, bei der Abstimmung der Evaluationsziele die teils divergierenden Sichtweisen der verschiedenen Stakeholder einzubeziehen und nach Möglichkeit für alle Gruppen tragbare Kompromisse zu finden. Damit dies geschehen kann, bedarf es zunächst der Identifizierung aller Beteiligten und Betroffenen, damit alle relevanten Perspektiven berücksichtigt werden können. Nach Beywl et al. (2007) sollte der Dialog mit den Beteiligten möglichst früh im Evaluationsprozess begonnen werden. Jedoch werden in vielen Fällen die entsprechenden Kontakte erst im Verlauf hergestellt, sodass die im Evaluationskonzept

beschriebenen Evaluationsziele ebenfalls zunächst als vorläufig zu betrachten sind.

Zur Qualitätssicherung von Evaluationen wurden Evaluationsstandards entwickelt (vgl. Abschnitt 2.2). Bereits in der Konzeptionsphase sollten Evaluatoren überlegen, wie diese Richtlinien im Evaluationsverlauf eingehalten werden können und welche Maßnahmen zu diesem Zweck zu treffen sind. Das Evaluationskonzept und das Angebot sollten unter Berücksichtigung der Evaluationsstandards erstellt werden.

Wird das Angebot der Evaluatoren vom Auftraggeber angenommen (bzw. im entsprechenden Fall umgekehrt), kommt es im nächsten Schritt zum Vertragsabschluss. Im Evaluationsvertrag sollten alle Rechte und Pflichten von Auftragnehmer und Auftraggeber festgehalten werden, darunter z. B. terminliche Vereinbarungen oder der Zugang zu Informationen. Auch während der Projektlaufzeit empfiehlt es sich, Abweichungen von diesen Übereinkünften vertraglich niederzulegen.

Planungsphase. Ist die vertragliche Situation geklärt, gilt es in der nun folgenden Phase, das Evaluationsvorhaben im Detail zu planen. Das Evaluationsteam entwickelt konkrete Fragestellungen und Hypothesen, welche sich zumindest teilweise an den Vorgaben des Auftraggebers orientieren. Zudem sollten die Evaluatoren darauf achten, dass die Evaluationsfragestellungen nicht einseitig ausgelegt sind, sondern alle für eine umfassende Bewertung des Evaluandums erforderlichen Faktoren Beachtung finden. Da zu evaluierende Maßnahmen eher selten in einen fundierten theoretischen Hintergrund eingebettet sind, können auch die Hypothesen meist nicht theoriegeleitet formuliert werden, sondern werden nach pragmatischen Gesichtspunkten entwickelt (für Informationen zu eher theoriebasierten Evaluationsansätzen vgl. Döring, 2009).

Weiterhin erfolgt nun die Untersuchungsplanung, welche die Stichprobe, das Studiendesign und die Erhebungsinstrumente umfasst. Bezüglich der Stichprobenzusammensetzung ist zu beachten, dass in der Evaluationspraxis nur selten Zufallsstichproben gezogen werden können, sondern häufig Ad-hoc-Stichproben (z. B. Anwesende bei einer Veranstaltung) oder das Schneeballprinzip zur Anwendung kommen. Deren Auswirkungen auf die Repräsentativität der Stichprobe und in der Folge auf die Gültigkeit der Ergebnisse sind zu berücksichtigen. Bei Evaluationen sind zudem nach Möglichkeit auch

solche Personen einzubeziehen, die an der Maßnahme nicht (mehr) teilnehmen, um Informationen über Nutzungsbarrieren oder Unzufriedenheit zu erhalten.

Beim Studiendesign kann unterschieden werden zwischen Feld- und Laboruntersuchungen, wobei Evaluationen in der Regel in Form von Feldstudien stattfinden. Hinsichtlich der Datenerhebung besteht die Wahl zwischen einem quantitativen und einem qualitativen Ansatz. Dabei stellt derzeit die Kombination beider Ansätze im Rahmen eines Mixed-Methods-Vorgehens das Mittel der Wahl dar (vgl. Stockmann, 2004). Weiterhin gibt es Quer- und Längsschnittstudien, wobei letztere mehrere Erhebungen im Zeitverlauf enthalten (z. B. Prä- und Posterhebung, also Erhebungen vor und nach der Intervention; ggf. Follow-Up-Erhebung zur Untersuchung langfristiger Effekte). Zudem können Kontrollgruppen einbezogen werden. Dabei ist z. B. zu überlegen, ob eine echte experimentelle Studie, d. h. Zuordnung der Probanden zu Versuchs- und Kontrollgruppe per Zufall (randomisiert), realisierbar ist. Dieses Design schließt von vornherein viele Validitätsgefährdungen aus und ist daher erstrebenswert, in der Praxis jedoch gerade bei Evaluationsstudien schwierig umzusetzen, da die zu untersuchenden Gruppen in der Regel bereits eingeteilt sind (z. B. Vergleich von Schulklassen). Evaluationsstudien werden daher häufig in quasiexperimenteller Form durchgeführt, wobei statt einer Randomisierung die Verwendung vorhandener Gruppen erfolgt und deren Vergleichbarkeit durch die Kontrolle relevanter Parameter zu erreichen versucht wird. Es gibt aber auch viele Evaluationen, die mit schwachen Designs durchgeführt werden, indem z. B. nach der Durchführung eines Kurses die Teilnehmer freiwillig einen Fragebogen zu ihrer Zufriedenheit mit dem Kurs ausfüllen. Hier wäre aufgrund der freiwilligen Teilnahme nicht nur die Repräsentativität der Ergebnisse fraglich, sondern darüber hinaus auch die Validität von daraus abgeleiteten Aussagen zur allgemeinen Qualität des Kurses, wenn nicht zusätzlich noch beispielsweise Verhaltensbeobachtungen erhoben werden. Natürlich zeichnen sich reine Fragebogenerhebungen durch ihre Ökonomie aus und können daher leicht angewandt werden. Diese Bemühungen zur Qualitätssicherung sind auch grundsätzlich lobenswert und solch kleinere Untersuchungen können, gemeinsam mit methodisch ausgereifteren Studien, durchaus zum Erkenntnisgewinn beitragen (vgl. Lösel, Schmucker, Plankensteiner & Weiss, 2006).

Da eine detailliertere Besprechung der Studiendesigns, der verschiedenen Validitätsgefährdungen und der Stichprobenthematik den Rahmen der vorliegenden Arbeit überschreiten würde, wird an dieser Stelle auf die in Fülle vorhandene Literatur zu diesem Bereich verwiesen (z. B. Beywl et al., 2007; Bortz & Döring, 2006; Holling, 2009; Kury & Obergfell-Fuchs, 2012; Lösel, Köferl & Weber, 1987; Meyer, 2007; Rossi et al., 1988; Rost, 2000; Wittmann, 1985; Wottawa & Thierau, 1998).

Um die Evaluationshypothesen beantworten zu können, bedarf es einer darauf zugeschnittenen Datenerhebung. Die Forscher müssen sich überlegen, wie die interessierenden Merkmale operationalisiert und mit welchen Methoden sie gemessen werden können. Zur Auswahl steht prinzipiell das gesamte Spektrum der sozialwissenschaftlichen Methoden, darunter z. B. Befragungen, Testverfahren und Verhaltensbeobachtung. Für viele Fragestellungen existieren bereits entsprechende Instrumente, beispielsweise Persönlichkeitsfragebögen oder Kategoriensysteme für Verhaltensbeobachtungen. Sofern dies möglich ist, sollten erprobte Verfahren zur Anwendung kommen, da deren Gütekriterien in der Regel geprüft sind und sie aufgrund größerer Bekanntheit auch eine bessere Vergleichbarkeit mit anderen Studien ermöglichen. Da in Evaluationen jedoch oftmals speziellen Fragen nachgegangen wird, ist es in diesen Fällen häufig notwendig, vorhandene Verfahren zu adaptieren oder neue Instrumente zu entwickeln. Deren Qualität sollte zunächst überprüft werden, was einen gewissen Aufwand bedeutet. Insgesamt ist dafür Sorge zu tragen, dass die Erhebungen keine unangemessen große Belastung für die Probanden darstellen, weshalb der Erhebungsumfang möglichst gering zu halten ist.

Als letzter Punkt in der Planungsphase ist die Projektplanung zu nennen, welche auf die Umsetzung des erarbeiteten Untersuchungsplans in der Praxis abzielt. Es wird ein zum Untersuchungsplan passender Zeitplan entwickelt bzw. die Planungen zu Zeit, Kosten und Aufgaben aus dem Evaluationskonzept werden angepasst.

Realisierungsphase. Die Realisierungsphase umfasst die Erhebung, Auswertung und Interpretation der Evaluationsdaten sowie das Projektcontrolling. Da die Datenerhebung häufig im Feld stattfindet, ist besonders darauf zu achten, dass Störungen der Praxis gering gehalten werden. Evaluatoren müssen den Datenschutz gewährleisten und die sozialen Normen und rechtlichen Regelungen des jeweiligen Bereichs beachten. Zudem ist ihre Flexibilität gefragt, da sich die

Rahmenbedingungen im Feld ändern können und Untersuchungsdesign und Instrumente dann angepasst werden müssen. Die erhobenen Daten sind mit angemessenen quantitativen bzw. qualitativen Auswertungsmethoden zu analysieren, wobei wie bei allen Forschungsprojekten die gängigen Voraussetzungen für die Anwendung quantitativer Verfahren zu prüfen sind. Generell ergeben sich bei Evaluationsstudien im Feld häufig schwache Effekte, die nur bei sehr großen Stichprobenzahlen signifikant würden. Daher sind neben signifikanten Ergebnissen auch Tendenzen interessant. Auch wenn mehrere Ergebnisse in die gleiche Richtung weisen, kann dies selbst bei fehlender Signifikanz einen wichtigen Hinweis bedeuten. Bei der Interpretation der Daten schließlich geht es darum, die Ergebnisse zu deuten und daraus ggf. Empfehlungen zu ziehen. Erwähnenswert ist, dass in anderen Forschungsbereichen gegensätzliche Ergebnisse diskutiert und daraus offene Fragen für künftige Forschung abgeleitet werden können. Evaluierende haben dagegen die selten leichte Aufgabe, auch bei divergierenden Befunden eine Synthese herzustellen und ein schlüssiges Gesamtergebnis zu erarbeiten. Die Interpretation sowie auch die Datenanalyse erfolgen bei summativen Evaluationen am Ende und sind auf zusammenfassende Aussagen zur Wirksamkeit hin ausgerichtet. Bei formativen Evaluationen werden hingegen Zwischenergebnisse ausgewertet und interpretiert, um frühzeitig Empfehlungen zur Optimierung des Evaluandums liefern zu können.

Das Projektcontrolling beinhaltet die Überwachung der Arbeitsvorgänge und den Abgleich mit den Plänen und Zielen, sodass Hindernisse bei der Projektdurchführung erkannt und Lösungen erarbeitet werden können. Dieser Bereich ist in der Realisierungsphase besonders relevant, spielt aber während des gesamten Evaluationsprozesses eine Rolle.

Abschlussphase. In der letzten Evaluationsphase werden die Ergebnisse den Beteiligten vermittelt. Dies geschieht üblicherweise in Form eines umfassenden Endberichtes, der den Evaluationsprozess ausführlich dokumentiert und neben den Ergebnissen auch die Beantwortung der Hypothesen sowie Schlussfolgerungen und ggf. Empfehlungen enthält. Bei formativen Evaluationen werden zudem (in früheren Evaluationsphasen) kürzere Zwischenberichte abgeliefert, die je nach Absprache auch eher informell z. B. anhand von E-Mails oder bei Treffen kommuniziert werden können. Wichtig ist, dass bei der (schriftlichen

und mündlichen) Berichterstattung berücksichtigt wird, dass die Ergebnisse der jeweiligen Zielgruppe verständlich sein müssen, weshalb z. B. spezifische Fachausdrücke zu vermeiden oder zumindest zu definieren sind.

Da allein durch die Vorlage eines Abschlussberichtes nicht sichergestellt ist, dass die darin präsentierten Ergebnisse auch zu einer Umsetzung in die Praxis führen, sollte die Nutzung der Ergebnisse unterstützt werden, indem verschiedenen Beteiligten die für sie jeweils relevanten Ergebnisse und Handlungsempfehlungen zusammengestellt und übermittelt werden. Dies kann schriftlich, beispielsweise durch Erstellung eines Merkblatts geschehen, aber auch persönlich, z. B. in Form von Workshops. Da die Nutzung von Evaluationsergebnissen ein wichtiges und nicht unumstrittenes Thema darstellt, wird in Abschnitt 2.1.6 nochmals ausführlicher darauf eingegangen.

Es wird empfohlen, dass auch Evaluationsstudien evaluiert, also sogenannte Meta-Evaluationen durchgeführt werden. Solche Bewertungen von Evaluationen können intern erfolgen und haben dann meist zum Ziel, den Evaluationsprozess systematisch zu prüfen und mögliche Schwierigkeiten aufzudecken, um durch Lernprozesse die Qualität von Evaluationen zu erhöhen. Doch auch bei Unstimmigkeiten bezüglich der Qualität einer Evaluationsstudie, wie sie im Nachhinein z. B. zwischen Auftraggeber und Evaluationsteam vorkommen können, kann eine Bewertung der Evaluation nützlich sein und wird in solchen Fällen in der Regel von externen Meta-Evaluatoren vorgenommen. Die Thematik der Meta-Evaluation wird in Abschnitt 2.2 intensiver behandelt.

Evaluationsergebnisse werden nicht immer publiziert, wie dies sonst im Forschungsbereich üblich ist. Das hat verschiedene Gründe. So ist teilweise die Thematik einer Evaluation zu spezifisch, teils willigen Auftraggeber nicht in eine Veröffentlichung der Ergebnisse ein, und auch Zeitmangel auf Seiten der Evaluierenden sowie datenschutzrechtliche Bedenken können eine Rolle spielen. Doch ist eine Publikation von Evaluationsstudien und auch von Meta-Evaluationen grundsätzlich anzustreben, damit Evaluation als Forschungsgebiet anerkannt bleibt und sich weiterentwickeln kann. In diesem Zusammenhang weisen Spiel und Kollegen (Spiel et al., 2012) auch darauf hin, dass Lernergebnisse und Erfahrungen aus vergangenen Evaluationen bisher kaum für die Optimierung künftiger Projekte nutzbar gemacht werden (vgl. Abschnitt 2.2.5.1).

Zusammengenommen kann festgehalten werden, dass Evaluationen in der Regel fünf Phasen (Initiierung, Konzeption, Planung, Realisierung, Abschluss) umfassen, welche jeweils spezifische Aufgaben mit sich bringen. Nach diesem Überblick über die Schritte eines Evaluationsprozesses werden im nächsten Abschnitt Bedeutung und Grenzen von Evaluationen erläutert. Danach wird die Thematik der Nutzung von Evaluationsergebnissen aufgegriffen und vertieft und anschließend die Rolle der Evaluationsforschung im Spannungsfeld zwischen Wissenschaft und Praxis diskutiert.

2.1.5 Bedeutung und Grenzen von Evaluationen

Wie bereits erwähnt wurde, erfreuen sich Evaluationen in Deutschland einer wachsenden Nachfrage (vgl. z. B. DeGEval, 2007; Stockmann, 2002a), weshalb an dieser Stelle die Bedeutung von Evaluationen erörtert, aber auch auf ihre Grenzen eingegangen wird.

Evaluationen werden in der Regel eingesetzt, um systematische Informationen über eine Maßnahme, z. B. ein Interventionsprogramm zu erhalten. Diese können dazu dienen, das Programm zu optimieren (formative Evaluation) und/oder die Grundlage für Entscheidungen bezüglich der Fortführung oder Beendigung des Programmes zu liefern, indem die Programmwirksamkeit untersucht wird (summative Evaluation). Gerade in Zeiten knapper werdender Ressourcen gewinnt der Nachweis der Wirksamkeit von Programmen an Bedeutung, da keine Mittel für unwirksame Maßnahmen verschwendet werden sollen (Stockmann, 2002a). In der Medizin beispielsweise ist dieses Vorgehen der „evidenzbasierten Praxis“ bereits etabliert (Kuhn, Lampert & Ziese, 2012; Kury & Obergfell-Fuchs, 2012). So werden die Wirkungen (und Nebenwirkungen) von Medikamenten und anderen therapeutischen Maßnahmen seit langem überprüft, sodass diesbezüglich heute umfangreiche Erkenntnisse zur Verfügung stehen. Im Bereich sozialer Programme hingegen werden verhältnismäßig wenige Evaluationen durchgeführt (Lösel et al., 2006; Mittag & Hager, 2000), auch wenn in den letzten Jahren eine steigende Tendenz zu beobachten ist (Weiss, Schmucker & Lösel, 2015). Dies bringt mit sich, dass etliche Programme realisiert und finanziert werden, die mitunter keine (oder sogar negative) Auswirkungen auf die Teilnehmenden haben. Hier wäre eine Erhöhung der Evaluationsaktivitäten wünschenswert, denn auch wenn

Evaluationen zunächst einmal finanzielle Mehrausgaben bedeuten, können sie doch dazu führen, dass weniger unwirksame Programme angeboten bzw. Programme in Bezug auf ihre Wirksamkeit optimiert und so die Kosten langfristig gesenkt werden (Kury & Obergfell-Fuchs, 2012; Stockmann, 2010). Insgesamt leisten Evaluationen einen wichtigen Beitrag zu evidenzbasierter Politik und Praxis (DeGEval, n.d.b) und sind, wie Kury und Obergfell-Fuchs konstatieren, „für die Fortentwicklung von Programmen und Maßnahmen … unersetzlich“ (2012, S. 1548).

Gerade weil Evaluationen eine so hohe Bedeutung im Rahmen der evidenzbasierten Praxis beizumessen ist, ist es wichtig, sich auch über ihre Grenzen im Klaren zu sein. Fitzpatrick et al. (2012) betonen, dass Evaluationsstudien nie eine genaue Abbildung der Realität liefern können, sogar wenn verschiedene Methoden zur Anwendung kommen. Dies liegt einerseits in den Grenzen der Methoden (z. B. mangelnde Generalisierbarkeit) selbst begründet und beruht andererseits auf den Einschränkungen des Forschungsalltags, wie z. B. dem Budget, den Kompetenzen des Evaluationsteams und der Beschaffenheit des Kontextes. Weiterhin ist der politische Aspekt zu nennen, da Entscheidungsträgern vielfältige Informationsquellen zur Verfügung stehen, von denen Evaluationen nur eine darstellen. So gilt es z. B. auch verschiedene Interessen und Erfahrungen mit zu berücksichtigen, weswegen Evaluationen nur zu einem gewissen Teil Entscheidungen beeinflussen können. Und schließlich sind Evaluationen auch insofern eingeschränkt, als Evaluierende und andere Beteiligte nur an der jeweiligen einzelnen Evaluation interessiert sind und Evaluation nicht als übergreifendes Feld verstehen, welches die Entwicklung von Organisationen und Personen ermöglicht. Dieser Begrenzung kann entgegengewirkt werden, indem mehr Evaluationen durchgeführt und ihre Berichte bekannter gemacht werden, sodass z. B. auch ähnliche Interventionen von einer Programmevaluation profitieren können. Und auch Evaluatoren können ihre Fähigkeiten weiterentwickeln, wenn sie sich im Rahmen einer Meta-Evaluation kritisch prüfen (lassen).

2.1.6 Nutzen und Nutzung von Evaluationen

An eine weitere Grenze stoßen Evaluationen dort, wo ihre Ergebnisse nicht genutzt werden. Die gesamte Durchführung einer Evaluation erscheint fragwürdig, wenn sie keinen Nutzen nach sich zieht. Dennoch ist die Nutzung von Evaluationsergebnissen keineswegs selbstverständlich und stellt ein Thema dar, welches in der Forschungsliteratur vielfach diskutiert wurde und wird. Es wird daher im Folgenden näher beleuchtet.

2.1.6.1 Nutzenerwartungen

Zunächst ist zwischen dem intendierten Evaluationsnutzen und der tatsächlichen Nutzung zu unterscheiden (vgl. Widmer, 2008). Der intendierte Nutzen spiegelt sich in den Nutzenerwartungen der verschiedenen Beteiligten wider. Die Nutzenerwartungen von Auftraggebern beispielsweise decken sich üblicherweise weitgehend mit den unter Abschnitt 2.1.3.2 angeführten Evaluationsfunktionen, also dem Zweck, dem die Evaluation dienen soll. Dabei ist, wie bereits erwähnt, zu beachten, dass diese Nutzenerwartungen häufig nicht expliziert werden und ihre Explikation mitunter auch nicht gewollt ist, sofern z. B. strategische Ziele verfolgt und Evaluationen instrumentalisiert werden. Solche impliziten Nutzenerwartungen können den Nutzen des Evaluationsprojektes erheblich beeinträchtigen, wenn sie nicht expliziert werden, da die Evaluierenden die Studie auf andere Zielsetzungen hin ausrichten. Auch bei den von der Evaluation betroffenen Personen ist wenig Akzeptanz zu erwarten, wenn beispielsweise bei einem Interventionsprogramm, das vordergründig zum Zweck seiner Weiterentwicklung evaluiert wurde, nach Beendigung der Evaluation und unter Berufung auf ihre Ergebnisse Personal entlassen wird. Ein solches Vorgehen ist als ethisch höchst bedenklich zu bewerten und kann zudem negative Auswirkungen auf die Glaubwürdigkeit der Evaluatoren haben. Das Konfliktpotential durch (implizite) Nutzenerwartungen wird zudem noch dadurch erhöht, dass auch die anderen Stakeholder ihre eigenen, meist divergierenden Nutzenerwartungen besitzen und auch diese nur teilweise expliziert werden (DeGEval, 2008b; Gollwitzer & Jäger, 2014). In Evaluationsstandards wie denen der DeGEval wird daher gefordert, dass die Evaluationszwecke geklärt und die Werte der Beteiligten transparent gemacht werden (DeGEval, 2008b; vgl.

Abschnitt 2.2.3). Inwieweit dies in der Praxis umgesetzt werden kann, sodass auch implizite Nutzenerwartungen thematisiert und konstruktive Kompromisse gefunden werden, dürfte von der Kommunikationsfähigkeit und Akzeptanz der Evaluierenden abhängen (vgl. Abschnitt 2.1.7.2)

2.1.6.2 Nutzung von Evaluationsergebnissen

Das Ausmaß der tatsächlichen Verwendung von Evaluationsergebnissen ist umstritten. So haben etliche Autoren eine mangelhafte Nutzung festgestellt, die sich darin äußert, dass Evaluationsergebnisse nicht oder nicht in der vorhergesehenen Weise genutzt und teilweise sogar ignoriert werden (z. B. Spiel & Bergsmann, 2009; Stamm, 2003; Widmer, 2008). Während solche Befunde Zweifel an der Existenzberechtigung von Evaluationen wecken können, gibt es auch Studien, die eine gute oder zumindest teilweise Nutzung von Evaluationsergebnissen aufzeigen (z. B. Alexander, 2003; Keller-Ebert, 2006; Stamm, 2003). Brandt (2007) weist zudem darauf hin, dass die Frage der Nutzung nicht pauschal zu beantworten ist und unterscheidet folgende Nutzungsarten (vgl. auch Rossi et al., 1988; Widmer, 2008; für weitere Nutzungstypologien vgl. Stamm, 2003; Widmer, 2008):

1. **Direkter (instrumenteller) Nutzen**: Bei dieser Nutzungsart finden Evaluationsergebnisse unmittelbare Anwendung durch Auftraggeber und andere Beteiligte, indem z. B. auf ihrer Grundlage Entscheidungen getroffen oder Evaluationsempfehlungen in die Praxis umgesetzt werden.
2. **Konzeptioneller Nutzen**: Dieser liegt vor, wenn Evaluationen sich allgemein auf das Denken der Beteiligten auswirken, ohne direkte Konsequenzen nach sich zu ziehen. So kann beispielsweise nach der Evaluation einer Maßnahme keine Entscheidung fallen, während aber dennoch das Wissen über die Maßnahme vermehrt wurde.
3. **Überzeugungs- bzw. symbolischer Nutzen**: Hier werden Evaluationsergebnisse dazu verwendet, politische Standpunkte zu bestätigen oder diesen zu widersprechen. Dies kann z. B. der Fall sein, wenn bereits getroffene Entscheidungen nachträglich unter Berufung auf Evaluationen legitimiert werden. Evaluationen kommt somit die Funktion von Argumenten zu,

weshalb auch von „argumentativer“ Verwendung gesprochen wird.

4. **Prozessnutzen**: Diese Nutzungsart resultiert nicht, wie die drei erstgenannten, aus den Ergebnissen, sondern aus der Durchführung der Evaluation. Denn auch der Evaluationsprozess selbst kann Veränderungen bewirken, wie z. B. in der Denkweise der Beteiligten über den Evaluationsgegenstand oder auch in der Organisation der evaluierten Maßnahme.

Bei der Diskussion um die Verwendung von Evaluationsergebnissen erscheint eine zu starke Fokussierung auf die direkte Nutzung im Sinne klar erkennbarer Entscheidungen zu eng gefasst. Denn dass auf eine Evaluation keine Entscheidungen folgen, bedeutet nicht, dass sie keinen Nutzen erbracht hat. Gerade die konzeptionelle Nutzung bleibt meist weniger sichtbar, sollte deshalb jedoch nicht unbeachtet bleiben. Sie kann unterschiedliche wichtige Auswirkungen haben und z. B. Interesse für bestimmte Problemstellungen wecken oder auch künftige Programme beeinflussen (Brandt, 2007; Rossi et al., 1988). Und schließlich sollte sich der Pessimismus bezüglich der direkten Verwendung von Evaluationen auch deshalb in Grenzen halten, weil Evaluationen, wie unter Abschnitt 2.1.5 angesprochen, nur einen Teil des (politischen) Entscheidungsprozesses darstellen (Fabian, 2009; Rossi et al., 1988). Es wäre deshalb zu weit gegriffen, von jeder Evaluation zu erwarten, dass sie weitreichende Entscheidungen herbeiführt (zumal der Einfluss einer Evaluation auf eine getroffene Entscheidung nicht immer identifiziert werden, aber trotzdem vorhanden sein kann, wie Alexander, 2003, herausstellt). Evaluationen und ihre Nutzung sollten somit „nicht überbewertet[,] aber auch nicht vernachlässigt werden“ (Fabian, 2009, S. 501).

2.1.6.3 Einflüsse auf die Nutzung

Wie kann die Verwendung von Evaluationsergebnissen in der Praxis verbessert werden? Es hat sich gezeigt, dass sich einige Kriterien positiv auf die Nutzung auswirken (Balthasar, 2009; Brandt, 2007; Spiel & Bergsmann, 2009):

- Relevanz der Evaluation für Entscheidungsträger und/oder andere Beteiligte (die Bedeutung dieses Faktors wird als besonders hoch eingestuft)
- Einbeziehung der Stakeholder in Planung und Berichterstellung
- Reputation oder Glaubwürdigkeit der Evaluierenden
- Qualität der Ergebnisvermittlung (passender Zeitpunkt, angemessene Häufigkeit und verständliche Kommunikation)
- Unterstützende Prozesse zur Ergebnisnutzung oder Erarbeitung von Handlungsempfehlungen

Auf die Frage, wie Evaluatoren zu einer besseren Nutzung beitragen können, wurden folgende Faktoren identifiziert (Brandt, 2007; Rossi et al., 1988; für weitere Hinweise zur Ausrichtung einer Evaluation auf Nutzen und Nutzung vgl. Beywl et al., 2007), die sich auch in den gängigen Evaluationsstandards wiederfinden (z. B. DeGEval, 2008b; Yarbrough, Shulha, Hopson & Caruthers, 2011; vgl. auch Abschnitt 2.2):

1. **Evaluierende müssen den Denkstil der Entscheidungsträger verstehen**: Dies ist wichtig, weil Evaluatoren und Entscheidungsträger in der Regel aus unterschiedlichen Praxisfeldern kommen und z. B. Politiker sich nicht in den Fachjargon der Wissenschaftler einarbeiten können, um den Evaluationsbericht zu verstehen. Es muss daher sichergestellt sein, dass die Ergebnisse in einer auf den Nutzer zugeschnittenen Weise vermittelt werden. Dazu ist es notwendig, dass Evaluatoren sich an die Denkweise der Entscheidenden anpassen.
2. **Evaluationsergebnisse müssen zeitnah erzielt und zum benötigten Zeitpunkt verfügbar gemacht werden**: Damit die Ergebnisse optimal genutzt werden können, müssen sie rechtzeitig gewonnen und kommuniziert werden, auch wenn dies mitunter mit Abstrichen bei der Gründlichkeit verbunden ist.
3. **Die unterschiedlichen Interessen und Sichtweisen der Beteiligten müssen berücksichtigt werden**: Je besser die verschiedenen Beteiligten sowohl in die Planung als auch in die Interpretation der Ergebnisse und die Berichterstellung

eingebunden werden und dadurch ihre Einstellungen Beachtung finden, umso mehr können und wollen sie die erzielten Ergebnisse nutzen.

4. **Die Nutzung und Dissemination der Ergebnisse sollten Bestandteil des Evaluationsdesigns sein**: Wenn Evaluatoren sich gezielt damit befassen, wie die Ergebnisse, aber auch die Vor- und Nachteile der durchgeführten Untersuchung den Nutzern nähergebracht werden können, erhöht dies die Nutzungschancen.
5. **Evaluationen sollten auch Nutzungskriterien beinhalten und deren Erreichung überprüfen**: Hierzu sollten sich Evaluierende und Auftraggebende auf Kriterien für eine erfolgreiche Umsetzung der Ergebnisse einigen und abschließend den Grad der Zielerreichung prüfen.

Neben diesen Anforderungen an Evaluatoren sollte natürlich auch bei den Beteiligten, insbesondere bei den Entscheidungsträgern, die Bereitschaft vorhanden sein, aus Evaluationsergebnissen zu lernen und diese zu nutzen (DeGEval, n.d.a; Wottawa & Thierau, 1998). Zudem sind Evaluationen nur dann nützlich, wenn sie professionell durchgeführt werden. Dazu wiederum bedarf es einerseits der fachlichen, methodischen und Sozialkompetenzen der Evaluatoren und andererseits ausreichender finanzieller Mittel, um qualitativ hochwertige Evaluationen realisieren zu können (DeGEval, n.d.b). Es gibt also etliche Einflüsse auf die Nutzung von Evaluationsergebnissen und viele Maßnahmen können ergriffen werden, um diese zu optimieren.

2.1.7 Evaluation zwischen Wissenschaft und Praxis

2.1.7.1 Evaluationsforschung und Grundlagenforschung

Es wurde bereits mehrfach darauf hingewiesen, dass wissenschaftliche Evaluationen zu den angewandten Forschungsdisziplinen zählen und Evaluatoren keine „reine“ Forschung im Sinne von Grundlagenforschung betreiben. Von letzterer sind Evaluationen in einigen Punkten abzugrenzen (vgl. Bortz & Döring, 2006; Fitzpatrick et al., 2012; Stockmann, 2007).

So ist Grundlagenforschung nicht auf einen bestimmten Zweck hin ausgerichtet, sondern dient dem Erkenntnisgewinn, der Ansammlung von Wissen. Evaluationen hingegen sind nicht zweckfrei. Sie werden in der Regel als Auftragsforschung durchgeführt, was bedeutet, dass die Ziele und Fragestellungen vom Auftraggeber vorgegeben werden und dadurch die Freiheit der Forschung eingegrenzt ist. Zwar bringen auftraggeberorientierte Evaluationen auch Erkenntnisgewinn mit sich, dieser stellt jedoch nicht das primäre Ziel dar (ebenda).

Ein anderer Unterschied besteht darin, dass die prinzipielle Wertfreiheit der Forschung bei Evaluationsstudien per definitionem nicht eingehalten werden kann, denn Evaluationen sollen bewerten und zu bestimmten Fragen Stellung nehmen. Dieser Aspekt ist als relativ unproblematisch zu betrachten, sofern Evaluierende ihre Beurteilungen nach bestem Gewissen und vor allem auf der Grundlage sorgfältig erhobener und ausgewerteter Daten abgeben. Mitunter wird jedoch der Vorwurf erhoben, dass Evaluationen zu unkritisch und von vornherein so gestaltet seien, dass die erwünschten Ergebnisse mit einer hohen Wahrscheinlichkeit erzielt werden (Bortz & Döring, 2006; Roth, 2012). Solche Fälle, in denen sich Forscher instrumentalisieren lassen, kommen in der Praxis sicherlich vor und sollten nicht verschwiegen werden. Es sollte hieraus jedoch kein generelles, abwertendes Urteil über Evaluationen gefällt werden. Im Übrigen können immer auch die Erfahrungen, Vorurteile und Eigeninteressen der Forscher eine Rolle spielen und sich beispielsweise auf die Ergebnisse und die Berichterstattung auswirken. Solche Einflüsse sind sowohl bei der Evaluations- als auch bei der Grundlagenforschung möglich. An beide Forschergruppen wird daher der Appell gerichtet, sich der Wissenschaftsethik entsprechend zu verhalten und einem möglichen Bias entgegenzuwirken. Zudem existieren mittlerweile Evaluationsstandards (vgl. Abschnitt 2.2), die zur Qualitätssicherung beitragen und an denen Evaluatoren ihre Arbeit orientieren können (Bortz & Döring, 2006; Fitzpatrick et al., 2012; Stockmann, 2007).

Als weiteren Aspekt nennen Bortz und Döring (2006) die Publikation von Forschungsergebnissen. Es wurde bereits erwähnt, dass Evaluationsberichte aus verschiedenen Gründen nicht immer veröffentlicht werden (vgl. Abschnitt 2.1.4). Hier zeigt sich ein Gegensatz zur Grundlagenforschung, da die Publikation von Befunden dort Usus ist und explizit angestrebt wird. Ein Nachteil der geringeren Dissemination von Evaluationsberichten besteht darin, dass dadurch auch der Rezeption

und Kritik durch die Scientific Community von vornherein Grenzen gesetzt sind und somit diese wichtige Ressource zur Kontrolle und Verbesserung nicht optimal genutzt wird.

Ein weiterer wichtiger Unterschied beruht auf der Tatsache, dass Evaluationsforschung den Anforderungen der Praxis gerecht werden muss, um nützliche Ergebnisse zu erzielen. So ist z. B. eine Evaluation mit einem sehr anspruchsvollen Design, vielen Datenerhebungen im Längsschnitt und der Verwendung aller relevanten Erhebungsmethoden aus methodischer Sicht erstrebenswert, würde jedoch mehr Geld kosten, als Auftraggeber zu zahlen in der Lage sind. Zudem würde sie so viel Zeit in Anspruch nehmen, dass die Ergebnisse für Entscheidungen zu spät kämen, womit die Evaluation ihren Nutzen verfehlt hätte. Evaluatoren befinden sich stets in einem Spannungsfeld zwischen den Anforderungen der Wissenschaft und denen der Praxis (vgl. Erzberger & Kelle, 2014; Hornbostel, 2010; Rossi et al., 1988; Stockmann, 2007). Dieses Spannungsverhältnis kann auch als Relevanz-Validitäts-Dilemma bezeichnet werden: Eine Orientierung des Evaluierenden nur an den Erwartungen der Praxis führt zu relevanten, aber nicht sehr validen Aussagen, während eine Ausrichtung der Evaluation ausschließlich an wissenschaftlichen Maßstäben valide, aber kaum relevante Ergebnisse hervorbringt (Friede, 1987). Evaluatoren müssen daher ihre eigenen wissenschaftlichen Ansprüche reduzieren und akzeptieren, dass sie sich in der angewandten Forschung befinden und dass sie, um der Praxis von Nutzen zu sein, gewisse Abstriche bei wissenschaftlichen Standards machen müssen. Die Evaluationsforschung muss sich häufig auf ausreichend gute Designs und Methoden beschränken, während in der Grundlagenforschung meist die bestmögliche Methodik gewählt werden kann und sollte (Erzberger & Kelle, 2014; Moosbrugger & Schweizer, 2002; Rossi et al., 1988).

Andere Restriktionen der Evaluationsforschung im Vergleich mit der Grundlagenforschung bestehen z. B. in den unterschiedlichen Zeithorizonten von Forschern und Politikern bzw. Entscheidungsträgern (Rossi et al., 1988), in der Notwendigkeit einer zeitlichen Koppelung des Evaluationsprozesses an die Intervention bei Programmevaluationen (Bortz & Döring, 2006) sowie in der Festlegung der Zeit- und Finanzressourcen durch den Auftraggeber statt durch den Forscher selbst (Meyer, 2007). Rossi und Kollegen (1988) erwähnen außerdem, dass Grundlagenforscher in der Regel innerhalb einer einzigen Wissenschaftsdisziplin arbeiten, auf spezielle Forschungsgebiete

fokussieren und nur eine begrenzte Anzahl von Forschungsmethoden anwenden. Evaluationsforscher hingegen arbeiten häufig in unterschiedlichen Fachbereichen und müssen je nach Erfordernissen der jeweiligen Fragestellungen diverse Methoden kennen und anwenden können (vgl. auch Fitzpatrick et al., 2012). Dies bringt Anforderungen an die Ausbildung von Evaluatoren mit sich, welche in Abschnitt 2.1.7.3 erörtert werden.

Hinsichtlich des Ablaufs enthalten Evaluations- und Grundlagenforschung teilweise ähnliche Schritte, wie die Auswahl von Datenerhebungsmethoden, deren Anwendung und die Analyse und Interpretation der erhobenen Daten (Stockmann, 2007). Evaluationsstudien umfassen darüber hinaus aber noch weitere Aufgaben. So muss gemeinsam mit dem Auftraggeber ein Konsens bezüglich der Ziele erarbeitet werden, meist müssen die Ziele dafür konkretisiert und eingegrenzt werden. Es gilt einen Evaluationsvertrag abzuschließen und Kosten- sowie Zeitpläne aufzustellen. Zudem sind die Interessen der verschiedenen Beteiligten zu klären und Maßnahmen für eine Akzeptanz der Evaluation bei diesen Gruppen zu treffen (Holling, 2009). Der letztgenannte Punkt macht deutlich, dass Evaluationen immer in einem sozialen und auch politischen Kontext stehen, worauf nun näher eingegangen wird.

2.1.7.2 Evaluation als sozialer und politischer Prozess

Von Evaluationen sind in aller Regel verschiedene Personengruppen betroffen, die sogenannten *Stakeholder*. Sie finden somit stets in einem mitunter komplexen sozialen Umfeld statt (Holling, 2009). Evaluationsforschung ist zudem im weiten Sinne eine politische Tätigkeit, da sie sich mit Aspekten der Gestaltung und Steuerung der zu bewertenden Maßnahme befasst, indem sie z. B. Entscheidungsgrundlagen liefert, die Entwicklung und Planung von Programmen beeinflusst und/oder dazu beiträgt, Maßnahmen zu verbessern. Es ist wichtig, dass Evaluatoren sich der Bedeutung ihres Arbeitskontextes bewusst sind und ihre Evaluation auch als sozialen und politischen Prozess begreifen und entsprechend ausrichten (Brandt, 2007; Rossi et al., 1988). So ist z. B. für die Nutzung der Evaluationsergebnisse die Berücksichtigung der unterschiedlichen Beteiligten wichtig, wie in Abschnitt 2.1.6.3 erwähnt. Auch wurde bereits darauf hingewiesen, dass Evaluationen immer nur einen Beitrag zu politischen Entscheidungen

liefern können, nicht aber die alleinige Entscheidungsgrundlage darstellen und manchmal auch ignoriert werden (vgl. Abschnitte 2.1.5 und 2.1.6.3). Auch hier sollten Evaluatoren die Bedeutung ihrer Forschung angemessen einschätzen.

Wie viele und welche Beteiligten es gibt, ist bei jeder Evaluation unterschiedlich und zu Beginn von den Evaluierenden zu klären. Mögliche Stakeholder sind unter anderem die Auftraggeber der Evaluation, die Träger des Programms, die Programmmitarbeiter, die Zielgruppen und auch das Evaluationsteam selbst (Brandt, 2007; Rossi et al., 1988). Alle beteiligten Personengruppen haben eigene Perspektiven und Interessen an der Evaluation. So kann auf Seiten der Auftraggebenden der Wunsch nach raschen Ergebnissen zur Absicherung einer Entscheidung vorherrschen, es kann aber auch das Ziel sein, Zeit zu gewinnen, um eine Entscheidung hinauszuzögern. Die Programmmitarbeiter können sich eine Bestätigung ihrer Arbeit erhoffen, aber auch Angst vor möglichen negativen Konsequenzen der Evaluation haben. Die Evaluatoren können sich wissenschaftlichen Standards verpflichtet fühlen und sich wünschen, eine gute Studie zu realisieren. Zugleich können aber auch die Zielsetzung einer persönlichen Qualifikation oder der Wunsch nach Würdigung der eigenen Arbeit im Vordergrund stehen (Friede, 1987).

Aus der Tatsache, dass Evaluationen mit so vielfältigen Interessen und Erwartungen einhergehen, ergibt sich beinahe zwangsläufig ein gewisses Konfliktpotential. Denn Stakeholder, die eher positive Erwartungen an eine Evaluation haben, werden sich kooperativ verhalten, während bei Ängsten der Beteiligten mit Ablehnung und zum Teil auch mit Behinderung der Evaluationsdurchführung zu rechnen ist. Die Evaluationsergebnisse können, wenn sie den Positionen der Stakeholder widersprechen, negative Reaktionen bis hin zu nachträglichen Diskreditierungen der Evaluation oder der Evaluatoren hervorrufen. Auch während der Evaluationsdurchführung sind Beeinflussungsversuche des Evaluationsteams durch Stakeholder möglich (Brandt, 2007; Friede, 1987; Rossi et al., 1988). Während gegensätzliche Interessen bei Evaluationen somit ein zentrales Thema ausmachen, werden diesbezügliche Erfahrungen und die daraus insbesondere für Evaluatoren entstehenden Dilemmata bislang kaum zum Gegenstand von Publikationen gemacht, obwohl Konflikte Entwicklungs- und Lernprozesse fördern können (Beywl, 2001).

Wie Brandt (2007) feststellt, ist es mitunter nicht möglich, die einander widersprechenden Interessen und die daraus entstehenden Konflikte vollständig aufzulösen und alle Beteiligten zufriedenzustellen. Evaluierende können nur ihr Bestes tun, indem sie alle Beteiligten von Beginn an in die Planung und Durchführung der Evaluation einbeziehen und ihre Erwartungen und Interessen soweit möglich explizit machen, um darauf eingehen zu können. Als besonders wichtig einzustufen ist dabei, dass Evaluatoren neben methodischen auch über soziale Fähigkeiten verfügen müssen (Brandt, 2007; Holling, 2009). Die beruflichen Anforderungen an Evaluatoren werden im Folgenden erörtert.

2.1.7.3 Berufsbild von Evaluatoren

Evaluatoren stellen in Deutschland keinen eigenständigen Berufsstand dar. So ist bisher der Marktzugang nicht einheitlich geregelt und es existieren so gut wie keine gezielten Aus- oder Weiterbildungsmöglichkeiten[2], um die für die Durchführung von Evaluationen notwendigen Kompetenzen zu erlangen. Dies liegt vor allem in der Transdisziplinarität der Evaluationsforschung begründet, da Evaluationen in verschiedensten Anwendungsfeldern durchgeführt werden und sich jeweils an den bereichsspezifischen Anforderungen orientieren müssen (Hense & Steckhan, 2014). Auch wenn Gemeinsamkeiten wie die Anwendung sozialwissenschaftlicher Methoden vorliegen, ist umstritten, ob Evaluationsforschung eine eigene Profession darstellen kann und sollte (Hornbostel, 2010). Zudem wird aktuell diskutiert, inwieweit standardisierte Ausbildungen vor dem Hintergrund der Heterogenität der Anwendungsfelder und entsprechenden Anforderungen möglich sind und wie solche Ausbildungsgänge gestaltet werden könnten[3] (Brandt, 2009; Hense & Steckhan, 2014).

Trotz der Vielgestaltigkeit von Evaluationen in der Praxis lassen sich einige Anforderungen beschreiben, die alle professionellen Evalutoren erfüllen sollten. Die DeGEval fasst diese in den folgenden vier

[2] Für einen Überblick zum aktuellen Stand der Aus- und Weiterbildung in Evaluation in Deutschland vgl. Hense und Steckhan (2014).

[3] Zu einem möglichen inhaltlichen und didaktischen Aufbau von Ausbildungsgängen hat die DeGEval in zwei Publikationen Empfehlungen erarbeitet, vgl. DeGEval (2008a) sowie Gutknecht-Gmeiner, Hense, Kihm, Lüth und Steckhan (2011).

Kompetenzfeldern zusammen, die Abbildung 3 veranschaulicht (DeGEval, 2008a; Hense & Steckhan, 2014).

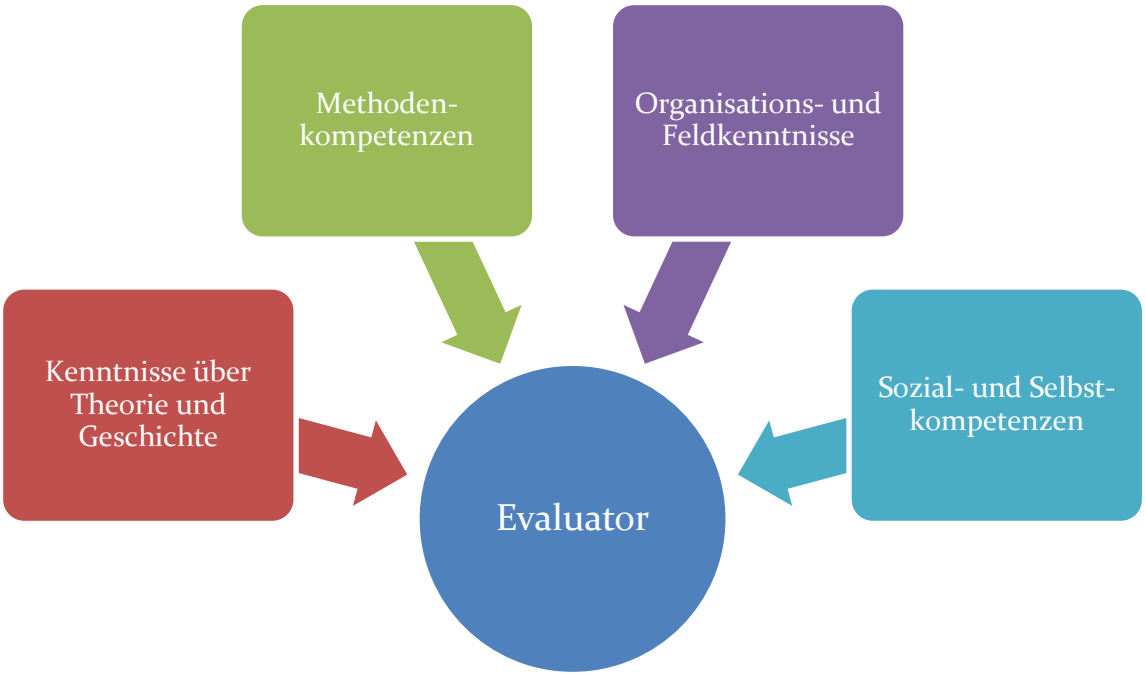


Abbildung 3: Anforderungen an die Kompetenzen von Evaluatoren (nach DeGEval, 2008a, S. 9)

1. **Kenntnisse über Theorie und Geschichte der Evaluation**: Evaluatoren müssen grundlegende Begriffe und Definitionen, aber auch den geschichtlichen Hintergrund der Evaluation kennen. Zudem ist die Kenntnis theoretischer und methodologischer Ansätze sowie von Evaluationsstandards notwendig, um Evaluationsfragestellungen einordnen und Evaluationen unter Berücksichtigung ihrer Möglichkeiten und Grenzen professionell durchführen zu können.
2. **Methodenkompetenzen**: Neben der Beherrschung quantitativer und qualitativer Forschungsmethoden werden hierunter auch Fertigkeiten im Rahmen der Projektführung verstanden, wie organisatorische und ökonomische Gesichtspunkte. Zudem bedeutet Methodenkompetenz, umfassendes Wissen über mögliche Untersuchungsdesigns und Erhebungsmethoden zu besitzen, um daraus die der Fragestellung angemessene Methodik auswählen zu können.

3. **Organisations- und Feldkenntnisse**: Die Organisationskenntnisse umfassen Wissen über die Merkmale, Grenzen und Funktionen von Organisationen, während zu den Feldkenntnissen einerseits Rechts- und Verwaltungswissen und andererseits spezifische Kenntnisse der jeweiligen Anwendungsgebiete zählen.
4. **Sozial- und Selbstkompetenzen**: Da in Evaluationen stets mit verschiedenen Stakeholdern zusammengearbeitet wird, sind die für eine positive Gestaltung der persönlichen Kontakte notwendigen sozialen sowie Kommunikations- und Kooperationskompetenzen eine weitere Anforderung an Evaluatoren. Auch die Fähigkeiten zum Selbstmanagement, zur Selbstreflexion und zur Problemlösung fallen darunter.

Da die DeGEval diese vier Kompetenzfelder im Hinblick auf eine Berücksichtigung in der Aus- und Weiterbildung erarbeitet hat, werden sie noch durch einen fünften Aspekt, **Praxis der Evaluation**, ergänzt. Dieser bildet jedoch keine Kompetenzen von Evaluierenden ab, sondern vielmehr die Forderung, dass die Ausbildung von Evaluatoren Praxiseinblicke umfassen soll sowie den Hinweis, dass viele der in den vier Kompetenzbereichen beschriebenen Fähigkeiten nur in der Praxis erlernt und gefestigt werden können (DeGEval, 2008a).

Die von der DeGEval dargestellten vier Kompetenzfelder sind als umfassend zu bewerten, schließen sie doch die in der weiteren Literatur genannten Anforderungen an Evaluatoren ein (vgl. Bortz & Döring, 2006; Holling, 2009; Wottawa & Thierau, 1998). Eine Betrachtung der hier erläuterten Anforderungen an Evaluierende macht zudem deutlich, dass diese in ihrer Gesamtheit nur selten in einer einzelnen Person vereinigt werden können. Daher empfiehlt sich die Bildung von Evaluationsteams, die sich in ihren Kompetenzen gegenseitig ergänzen (Brandt, 2007).

2.2 Evaluationsstandards und Meta-Evaluation

Nachdem im vorigen Kapitel eine Einführung in die Evaluationstheorie gegeben wurde, wird im Folgenden dargelegt, welche Maßnahmen zur Qualitätssicherung von Evaluationen unternommen werden können. Der Fokus liegt dabei, übereinstimmend mit dem Thema der vorliegenden Arbeit, auf Evaluationsstandards und Meta-Evaluation. Nach einem einführenden Abschnitt zu Evaluationsstandards werden verschiedene Richtlinien vorgestellt und anschließend die deutschen Standards für Evaluation im Detail besprochen. Es folgen eine Einführung in das Themengebiet der Meta-Evaluation und schließlich eine Darstellung des Forschungsstandes zu Evaluationsstandards und Meta-Evaluationen.

2.2.1 Evaluationsstandards – Einführung

Vor dem Hintergrund einer stetigen Zunahme an Evaluationsaktivitäten (vgl. Abschnitt 2.1.2) stellt sich die Frage nach einem Maßstab zur Bewertung von Evaluationen, um deren Qualität beurteilen zu können (Atria et al., 2006; DeGEval, 2008b). Zu diesem Zweck wurden spezielle Bewertungskriterien, sogenannte *Evaluationsstandards* bzw. *Standards für Evaluation* entwickelt. Angesichts des bereits geschilderten Entwicklungsvorsprungs der Evaluationsforschung in den USA gegenüber Deutschland und Europa (Abschnitt 2.1.2) verwundert es nicht, dass auch Evaluationsstandards zuerst von amerikanischen Autoren veröffentlicht und in der Folge unter anderem im deutschsprachigen Raum adaptiert wurden. Mittlerweile existiert eine Reihe verschiedener Kriterienkataloge, welche im nächsten Abschnitt überblicksartig dargestellt werden. Diese Richtlinien weisen untereinander viele Überschneidungspunkte auf und fordern im Allgemeinen, dass bei Evaluationsprojekten die folgenden Aspekte berücksichtigt werden (DeGEval, 2008b; Meyer, 2002; Moosbrugger & Schweizer, 2002; Yarbrough et al., 2011):

- **Nützlichkeit**: Evaluationen sollen nützlich sein, d. h. den Evaluationszweck erfüllen und dem Informationsbedarf der Beteiligten entsprechend angelegt werden.

- **Durchführbarkeit**: Hierunter werden Empfehlungen zum Umgang mit Rahmenbedingungen verstanden, sodass Evaluationen realistisch geplant und durchgeführt werden.
- **Fairness**: Evaluatoren sollen sich ethischen Grundsätzen verpflichten und einerseits respektvollen Umgang mit den Stakeholdern pflegen sowie andererseits der Forschungsethik gemäß handeln.
- **Genauigkeit**: Es wird eine Reihe von Empfehlungen gegeben, damit Evaluationen fundierte und ausführliche Informationen erbringen.

Es sollte an dieser Stelle erwähnt werden, dass die Einhaltung von Evaluationsstandards nicht obligatorisch ist. Einerseits gibt es keine Zertifizierung für Evaluatoren, welche sich solchen Richtlinien verpflichten, andererseits zieht eine Nicht-Befolgung in der Regel keine Sanktionen nach sich (Beywl & Widmer, 2006). Es gibt auch einige kritische Stimmen, die sich darauf beziehen, dass das Feld der Evaluationsforschung zu heterogen ist, als dass einheitliche Richtlinien sinnvoll wären, und dass solche Standards die Kreativität einschränken können (Beywl & Speer, 2004; Love & Russon, 2005; Stufflebeam, 2004). Dennoch empfiehlt es sich, dass Evaluierende ihrer Arbeit gängige Evaluationsstandards zugrunde legen, denn diese dienen dem Zweck der Qualitätssicherung und -entwicklung von Evaluationen (vgl. Beywl, 2001; Bortz & Döring, 2006; DeGEval, 2008b; Yarbrough et al., 2011):

So bieten sie Evaluierenden in jeder Phase einer Evaluation wichtige Orientierungshilfen und Entscheidungsgrundlagen. Sie stellen zudem ein Instrument dar, welches den Austausch über die Qualität von Evaluationen fördert, indem es einen einheitlichen Bezugsrahmen liefert. Auch für die Kommunikation zwischen Auftraggebenden und Evaluatoren können die Standards nützlich sein, da Evaluierende den Auftraggebern anhand der Standards verdeutlichen können, was die Evaluation leisten kann und was nicht, so dass in der Folge die Auftraggebenden wissen, welche Erwartungen und Ansprüche sie stellen können. Durch die Beachtung der Standards können die Interessen aller an der Evaluation Beteiligten, einschließlich der Evaluierenden, besser gewahrt werden. Den Standards kann auch eine konfliktpräventive Funktion zugeschrieben werden. Generell tragen Evaluationsstandards zu mehr Transparenz bei und ihre Einhaltung kann neben

der wissenschaftlichen Qualität auch den gesellschaftlichen Nutzen von Evaluationen erhöhen. Die Standards können zudem einen Beitrag zur Aus- und Weiterbildung im Evaluationsbereich liefern. Schließlich stellen sie eine wertvolle Grundlage für die Bewertung von Evaluationen dar, sowohl aus Sicht der Auftraggeber als auch seitens der Evaluierenden, die ihre eigene Arbeit kritisch betrachten und verbessern möchten (ebenda). Auf den Aspekt der Evaluationsbewertung wird in Abschnitt 2.2.4 näher eingegangen.

2.2.2 Standards und Richtlinien für Evaluationen – ein Überblick

Im Folgenden werden einige Evaluationsrichtlinien vorgestellt. Die Darstellung erhebt keinen Anspruch auf Vollständigkeit, sondern soll einen Überblick über internationale Werke zu Standards und Richtlinien für Evaluationen geben, während gleichzeitig diejenigen Werke etwas ausführlicher behandelt werden, die für die vorliegende Arbeit besonders relevant sind.

Zunächst sind die *Program Evaluation Standards* des Joint Committee on Standards for Educational Evaluation (JCSEE) zu nennen. Diese amerikanischen Standards wurden erstmalig 1981 veröffentlicht. Nach einem Revisionsprozess erschien 1994 die zweite Ausgabe und nach erneuter Überarbeitung liegt aktuell die dritte Auflage aus dem Jahr 2011 vor (Yarbrough et al., 2011). Dem JCSEE kann durchaus Pionierarbeit auf dem Gebiet der Evaluationsstandards bescheinigt und seine Program Evaluation Standards können als Standardwerk auf diesem Gebiet bezeichnet werden, auf dem viele andere Richtlinien basieren. Die Richtlinien des JCSEE gelten für Programmevaluationen im Bildungskontext in den USA und Kanada. Sie sind in die vier Bereiche *Utility*, *Feasibility*, *Propriety* und *Accuracy* untergliedert, welche jeweils zwischen vier und acht Einzelstandards umfassen. In der aktuellen dritten Auflage wurde zudem *Evaluation Accountability* als neuer, fünfter Bereich ergänzt, welchem drei Einzelstandards zugeordnet sind. Die Inhalte der amerikanischen Standards decken sich weitgehend mit denen der DeGEval. Da letztere im Abschnitt 2.2.3 detailliert beschrieben werden, wird auf eine nähere Ausführung der JCSEE-Standards an dieser Stelle verzichtet.

Die Arbeit des JCSEE hatte maßgeblichen Einfluss auf die Entwicklung von Evaluationsstandards in Europa. So weisen die Standards der Schweiz und Deutschlands eine explizite Orientierung an und Überschneidung mit ihrem amerikanischen Pendant auf. Als erste Evaluationsgesellschaft in Europa veröffentlichte die Schweizerische Evaluationsgesellschaft (SEVAL) im Jahr 2000 eigene Evaluationsrichtlinien, die sogenannten *SEVAL-Standards* (Widmer, Landert & Bachmann, 2000; vgl. auch Widmer, 2005). Diese unterteilen sich, ebenso wie die JCSEE-Standards, in die vier Gruppen *Nützlichkeit*, *Durchführbarkeit*, *Korrektheit* und *Genauigkeit*, die sich jeweils aus 3 bis 10 Einzelstandards zusammensetzen.

Die deutschen Evaluationsstandards, die *Standards für Evaluation* (DeGEval, 2008b) wurden von der DeGEval erarbeitet. Der Entstehungsprozess begann im Jahr 2000 mit einer Mitgliederbefragung, welche in dem Beschluss mündete, eigene Standards nach dem Vorbild der amerikanischen und der schweizerischen Richtlinien zu entwerfen. Dabei wurde explizit keine Neuformulierung oder Umstrukturierung vorgenommen, um die Vergleichbarkeit mit den anderen Werken zu erhalten und so den internationalen Dialog über die Standards zu fördern (ebenda). Die deutschen Standards, die 2001 verabschiedet wurden und aktuell in der vierten, unveränderten Auflage von 2008 vorliegen, beinhalten demnach analog zu den SEVAL-Standards die vier Kategorien *Nützlichkeit*, *Durchführbarkeit*, *Fairness* und *Genauigkeit*, welche wiederum jeweils zwischen drei und neun Einzelstandards umfassen. Im nächsten Abschnitt werden die DeGEval-Standards ausführlich behandelt.

Ein Werk, welches die Entwicklung der deutschen Richtlinien ebenfalls unterstützte, stellte die Übersetzung der JCSEE-Standards ins Deutsche dar, welche als „Handbuch der Evaluationsstandards“ 1999 erschien und in der aktuellen dritten Auflage um Informationen unter anderem zur Vergleichbarkeit und Übertragbarkeit der Standards ergänzt wurde (JCSEE, 2006). Auf dieses Werk wird in den DeGEval-Standards explizit verwiesen.

Es sei darauf hingewiesen, dass die Publikationen der DeGEval und der SEVAL sowie des JCSEE inhaltlich ähnlich sind, sich aber hinsichtlich ihrer Ausführlichkeit unterscheiden. So sind die deutschen und die schweizerischen Richtlinien vergleichsweise kurz gehalten und bestehen aus der Nennung der Standards sowie einer jeweils maximal halbseitigen Beschreibung. Die amerikanische Version hingegen

enthält deutlich ausführlichere Beschreibungen der Standards und umfasst zudem für jeden einzelnen Standard Empfehlungen zur Einhaltung, Hinweise auf häufige Fehler sowie Fallbeispiele zur Veranschaulichung.

Im Rahmen dieses Abschnitts können nicht alle Werke zum Thema „Evaluationsstandards" im Detail beschrieben werden. Jedoch werden an dieser Stelle noch einige weitere Richtlinien erwähnt, der interessierte Leser sei für weiterführende Informationen auf die Literaturangaben verwiesen.

So gibt Widmer (2005, 2006) einen Überblick über weitere Aktivitäten im Bereich der Evaluationsstandards in Europa. Der Forschungsstand in anderen Ländern und Kontinenten wird von Russon und Russon (2005) nachgezeichnet.

Ein einflussreiches Werk aus dem anglo-amerikanischen Raum stellen die *Guiding Principles for Evaluators* der American Evaluation Association dar. Diesen Richtlinien zufolge sollen Evaluierende ihre Arbeit an den folgenden fünf Prinzipien ausrichten: *Systematische Untersuchung* des Evaluandums, *Kompetenz* der Evaluatoren, *Rechtschaffenheit und Aufrichtigkeit* im Evaluationsprozess, *Respekt* allen Beteiligten gegenüber sowie *Verantwortungsgefühl* in Bezug auf das Gemeinwohl (American Evaluation Association, 2004; vgl. auch Meyer, 2002). Die Guiding Principles sind weniger detailliert als die JCSEE-Standards, gelten jedoch im Gegensatz zu diesen für alle Arten von Programmevaluationen, nicht nur für solche im Bildungskontext (Stufflebeam, 2001b).

Es wurden zudem Standards für spezielle Evaluationsbereiche entwickelt. So hat z. B. die DeGEval Richtlinien für den Bereich der Selbstevaluation herausgegeben (DeGEval, 2004c). Bei Selbstevaluationen sind die Evaluierenden zugleich Mitglieder oder Mitarbeiter der zu evaluierenden Maßnahme, bewerten also ihre eigene Arbeit (vgl. Abschnitt 2.1.3.4). Aus diesem Umstand resultieren einige Besonderheiten, vor allem bezüglich der Rollenverteilung und entsprechender spezieller Anforderungen. Die DeGEval rät dazu, dass die Standards für Evaluation auch bei Selbstevaluationen berücksichtigt werden sollen und gibt in der genannten Publikation Empfehlungen für deren Anwendung in diesem Bereich, die eine Gewichtung einzelner Aspekte beinhaltet und auf eine Anpassung an die besonderen Umstände bei Selbstevaluationen abzielt (ebenda).

Auch das JCSEE hat seine Program Evaluation Standards adaptiert und für Personalevaluationen die *Personnel Evaluation Standards* (JCSEE, 2009) sowie für Evaluationen im schulischen Bereich die *Student Evaluation Standards* (JCSEE, 2003) veröffentlicht. Inhaltlich unterscheiden sich diese beiden Richtlinien wenig von den Program Evaluation Standards, sondern stellen vielmehr Abwandlungen davon dar, die zum Zwecke der Anwendung in anderen Evaluationsfeldern vorgenommen wurden.

Abschließend soll auf Checklisten als Instrument zur praktischen Anwendung von Evaluationsstandards hingewiesen werden. Diese liegen bisher hauptsächlich für Richtlinien aus dem anglo-amerikanischen Raum vor und ihre Entwicklung und Etablierung geht insbesondere auf die Arbeit des amerikanischen Evaluationsforschers Daniel Stufflebeam zurück (z. B. Stufflebeam, 1999a; Stufflebeam, Goodyear, Marquart & Johnson, 2005). Die Checklisten können beispielsweise bei der Planung und Durchführung von Evaluationen eingesetzt werden, eine Unterstützung bei der Gestaltung von Verträgen darstellen oder bei der Bewertung von Evaluationen zur Anwendung kommen (Stufflebeam, 2001a; vgl. Abschnitt 2.2.5.2).

2.2.3 Die Standards für Evaluation der DeGEval

Im Folgenden werden die deutschen Evaluationsstandards der DeGEval detailliert erläutert, da sie eine wichtige methodische Grundlage der vorliegenden Arbeit darstellen. Dazu werden alle 25 Standards aufgeführt und mit den Kurzbeschreibungen aus der DeGEval-Publikation (DeGEval, 2008b) versehen[4]. Für zusätzliche Erläuterungen zu den einzelnen Standards sei auf ebendiese Publikation verwiesen, ebenso für einen differenzierten Vergleich mit den amerikanischen und schweizerischen Richtlinien. Der Abschnitt schließt mit einigen generellen Anmerkungen zu den Standards und ihrer Anwendung.

[4] Die in den Abschnitten 2.2.3.1 bis 2.2.3.4 angeführten Erläuterungen sind wörtlich den Standards für Evaluation der DeGEval (2008b, S. 23-38) entnommen, wobei Schreibfehler nicht übernommen wurden.

2.2.3.1 Nützlichkeitsstandards

Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.

Identifizierung der Beteiligten und Betroffenen (N1). Die am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.

Klärung der Evaluationszwecke (N2). Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, sodass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann.

Glaubwürdigkeit und Kompetenz des Evaluators/der Evaluatorin (N3). Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.

Auswahl und Umfang der Informationen (N4). Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellung zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen.

Transparenz von Werten (N5). Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind.

Vollständigkeit und Klarheit der Berichterstattung (N6). Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen, leicht zu verstehen und nachvollziehbar sein.

Rechtzeitigkeit der Evaluation (N7). Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehenden Entscheidungsprozessen bzw. Verbesserungsprozessen einfließen können.

Nutzung und Nutzen der Evaluation (N8). Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.

2.2.3.2 Durchführbarkeitsstandards

Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.

Angemessene Verfahren (D1). Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.

Diplomatisches Vorgehen (D2). Evaluationen sollen so geplant und durchgeführt werden, dass möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.

Effizienz von Evaluation (D3). Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.

2.2.3.3 Fairnessstandards

Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.

Formale Vereinbarungen (F1). Die Pflichten der Vertragsparteien einer Evaluation (was, wie, von wem, wann getan werden soll) sollen schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese neu auszuhandeln.

Schutz individueller Rechte (F2). Evaluationen sollen so geplant und durchgeführt werden, dass Sicherheit, Würde und Rechte der in eine Evaluation einbezogenen Personen geschützt sind.

Vollständige und faire Überprüfung (F3). Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst vollständig und fair überprüfen und darstellen, sodass die Stärken weiter ausgebaut und die Schwachpunkte behandelt werden können und die Rechte der in eine Evaluation einbezogenen Personen geschützt sind.

Unparteiische Durchführung und Berichterstattung (F4). Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation in Rechnung stellen. Berichte sollen ebenso wie der gesamte Evaluationsprozess die unparteiische Position des Evaluationsteams erkennen lassen. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden.

Offenlegung der Ergebnisse (F5). Die Evaluationsergebnisse sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.

2.2.3.4 Genauigkeitsstandards

Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

Beschreibung des Evaluationsgegenstandes (G1). Der Evaluationsgegenstand soll klar und genau beschrieben und dokumentiert werden, sodass er eindeutig identifiziert und so genau wie möglich zugänglich gemacht werden kann.

Kontextanalyse (G2). Der Kontext des Evaluationsgegenstandes soll ausreichend detailliert untersucht und analysiert werden.

Beschreibung von Zwecken und Vorgehen (G3). Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, sodass sie identifiziert und eingeschätzt werden können.

Angabe von Informationsquellen (G4). Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Information eingeschätzt werden kann.

Valide und reliable Informationen (G5). Die Verfahren zur Gewinnung von Daten sollen so gewählt oder entwickelt und dann eingesetzt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien quantitativer und qualitativer Sozialforschung orientieren.

Systematische Fehlerprüfung (G6). Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.

Analyse qualitativer und quantitativer Informationen (G7). Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation effektiv beantwortet werden können.

Begründete Schlussfolgerungen (G8). Die in einer Evaluation gezogenen Folgerungen sollen ausdrücklich begründet werden, damit die Adressatinnen und Adressaten diese einschätzen können.

Meta-Evaluation (G9). Um Meta-Evaluationen zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert und archiviert werden.

2.2.3.5 Hinweise zu den Standards und ihrer Anwendung

Die DeGEval-Standards verfolgen unter anderem die Zielsetzung, Evaluierenden in jeder Phase der Evaluation eine Orientierungshilfe zu bieten (die weiteren Zwecke wurden in Abschnitt 2.2.1 erläutert). Folglich können die einzelnen Standards den Phasen eines Evaluationsprozesses (vgl. Abschnitt 2.1.4) zugeordnet werden, wobei manche Standards in mehreren Phasen zur Anwendung kommen können. So ist beispielsweise die *Identifizierung der Beteiligten und Betroffenen (N1)* sowohl für die Initiierungs- und Konzeptions-, als auch für die Planungsphase relevant und spielt ebenso eine Rolle bei der Berichterstattung (Abschlussphase). Bei der Datenauswertung im Rahmen der Realisierungsphase sind vor allem (aber nicht nur) Genauigkeitsstandards wie z. B. *Analyse qualitativer und quantitativer Informationen (G7)* zu beachten. Eine umfassende Zuordnung der Standards zu Evaluationsphasen ist in den DeGEval-Standards enthalten (DeGEval, 2008b; vgl. auch Beywl et al., 2007; Döring, 2009; Schiffler & Hübner, 2000).

Bei der Anwendung der DeGEval-Standards ist zu berücksichtigen, dass es sich um Maximalstandards handelt. Das bedeutet, dass sie nicht bei jeder Evaluationsstudie eingehalten werden können oder müssen, sondern dass sie vielmehr einen Bezugsrahmen liefern, nach dem Evaluierende und andere Stakeholder sich richten können. Nicht

jeder Standard kann bei jeder Evaluation in der gleichen Weise erfüllt werden. Dies ist schon aufgrund der vielfältigen Rahmenbedingungen und Anforderungen, die sich in der Praxis ergeben, kaum möglich. Außerdem gibt es keine Angaben zur Priorisierung der Standards, sodass die Anwender die Gewichtung im konkreten Fall selbst vornehmen müssen. Hinzu kommt, dass manche Standards einander widersprechen. So werden z. B. Evaluatoren, die eine Studie auf höchstem wissenschaftlichem Niveau anstreben (und die Genauigkeitsstandards maximal erfüllen), feststellen, dass diese Studie zu viele Ressourcen bindet und den Durchführbarkeitsstandards nicht gerecht wird (DeGEval, 2008b; Meyer, 2002; Patry & Hager, 2000; Stockmann, 2007).

Den DeGEval-Standards zufolge sollen Evaluierende und Stakeholder in solchen Fällen einen der jeweiligen Situation angemessenen Kompromiss finden. Sofern einzelne Standards nicht eingehalten werden können, sollte dies kurz begründet werden. Generell sollte bei der Bewertung einer Evaluation nicht rein schematisch vorgegangen und lediglich beurteilt werden, wie viele Standards (nicht) eingehalten wurden. Vielmehr soll untersucht werden, „ob die anwendbaren Standards bei Planung und Durchführung der Evaluation berücksichtigt wurden und versucht wurde,[ ]ihnen im Rahmen der konkreten Bedingungen möglichst gerecht zu werden“ (DeGEval, 2008b, S. 19).

Widmer (1996) weist zudem darauf hin, dass die Evaluationsstandards die Kriterien der Unabhängigkeit und der Eindeutigkeit nicht erfüllen. So sind die einzelnen Standards nicht unabhängig voneinander, sondern vielmehr miteinander verwoben. Einige Richtlinien stehen mit anderen in einem positiven oder negativen Zusammenhang. Dies bedingt auch eine gewisse Uneindeutigkeit dahingehend, dass ein Sachverhalt für mehrere Standards relevant sein und daher nicht eindeutig nur einem Standard zugeordnet werden kann. Wenngleich Widmers Hinweise sich auf die Evaluationsstandards des JCSEE beziehen, dürften sie auch auf die deutschen Richtlinien übertragbar sein, da diese den amerikanischen sehr ähnlich sind.

2.2.4 Meta-Evaluation – Begriffsklärung und Einführung

Das Konzept der Meta-Evaluation soll nun definiert und näher beleuchtet werden. Der Begriff *Meta-Evaluation* (auch die Schreibweise

Metaevaluation ist möglich) wurde 1969 von Michael Scriven eingeführt (vgl. Stufflebeam, 1981, 2000) und meint die Evaluation einer Evaluation, also die Überprüfung der Qualität einer Evaluationsstudie (Döring, 2009; Stufflebeam, 1978, 1981, 2000, 2001b). Ergänzend zu dieser sehr knappen Begriffsbestimmung hat Stufflebeam eine anwendungsbezogenere Definition erarbeitet: „Metaevaluation is the process of delineating, obtaining, and applying descriptive information and judgmental information about the utility, feasibility, propriety, and accuracy of an evaluation in order to guide the evaluation and to publicly report its strengths and weaknesses" (Stufflebeam, 2000, S. 96). Aus dieser Definition wird der Bezug zu den Evaluationsstandards (Nützlichkeit, Durchführbarkeit, Fairness und Genauigkeit, vgl. Abschnitte 2.2.1 und 2.2.2) deutlich, da Evaluationen im Rahmen von Meta-Evaluationen auf deren Einhaltung hin geprüft bzw. bei deren Einhaltung unterstützt werden können. Zudem werden hier zwei Zwecke von Meta-Evaluationen benannt: Einerseits Evaluationen zu leiten, d. h. vor und/oder während des Evaluationsprozesses konstruktiv auf Verbesserungen hinzuwirken (formative oder auch proaktive Meta-Evaluation) und andererseits ihre Stärken und Schwächen öffentlich zu berichten, was üblicherweise erst nach Beendigung einer Evaluation geschieht (summative oder auch retroaktive Meta-Evaluation; vgl. Hanssen, Lawrenz & Dunet, 2008; Stufflebeam, 1981, 2001b).

Wie unter anderem Döring (2009) herausstellt, wird der Ausdruck „Meta-Evaluation" in der einschlägigen Literatur leider häufig synonym zur Bezeichnung „Meta-Analyse" verwendet, obwohl beide Begriffe inhaltlich unterschiedliche Konzepte bezeichnen. Erstere bezieht sich auf die Bewertung einer Evaluationsstudie, um die Qualität des Vorgehens und der erzielten Ergebnisse zu bestimmen, wohingegen Letztere die Synthese mehrerer Evaluationsstudien bedeutet und zum Ziel hat, deren Einzelbefunde systematisch zu einer zusammenfassenden Bewertung des Forschungsgegenstandes zu integrieren. Während also Meta-Analysen (auf der Grundlage mehrerer Studien) Aussagen über die Qualität einer Maßnahme treffen sollen, liegt der Zweck von Meta-Evaluationen darin begründet, die Qualität der Evaluationsforschung zu untersuchen und langfristig zu verbessern. Die oft synonyme Verwendung der beiden Termini könnte im Übrigen auch der Tatsache geschuldet sein, dass Meta-Evaluationen auch im Rahmen von Meta-Analysen angewendet werden, wenn es darum geht, nur Studien von ausreichender methodischer Qualität in die Synthese mit

aufzunehmen (vgl. auch Cooksy & Caracelli, 2005; Stufflebeam, 2001b). Um begrifflichen Missverständnissen vorzubeugen, wird an dieser Stelle ausdrücklich darauf hingewiesen, dass in vorliegender Arbeit mit Meta-Evaluation stets die *Bewertung einer Evaluation* im Sinne der oben genannten Definitionen gemeint ist.

Ähnlich wie bei Evaluationen (vgl. Abschnitt 2.1.3.4) kann auch bei Meta-Evaluationen zwischen einer internen und einer externen Durchführung unterschieden werden. Die interne Meta-Evaluation erfolgt durch das Evaluationsteam selbst und sollte zumindest in einer kurzen Form am Ende eines jeden Evaluationsprojektes realisiert werden. Externe Meta-Evaluationen werden von Außenstehenden durchgeführt und sind z. B. im Konfliktfall zwischen Auftraggebenden und Evaluierenden anzuraten (Döring, 2009; DeGEval, 2008b). Keine der beiden Formen kann pauschal als nützlicher eingestuft werden als die andere (Yarbrough et al., 2011). So ist es in der Praxis nicht immer möglich oder notwendig, externe Meta-Evaluatoren zu engagieren, sondern gerade formative Meta-Evaluationen können durchaus in angemessener Weise vom Evaluationsteam selbst realisiert werden (Stufflebeam, 2001b). Des Weiteren muss betont werden, dass interne Meta-Evaluationen nicht zwangsläufig weniger objektiv sind, auch wenn dies häufig unterstellt wird. Denn auch externe Meta-Evaluierende haben ihre Vorerfahrungen und Interessen. Daher können auch bei ihnen Interessenkonflikte auftreten, wenngleich diese bei internen Meta-Evaluatoren aufgrund deren Einbindung in die Evaluation und/oder das Programm als wahrscheinlicher gelten (Yarbrough et al., 2011).

Externe Meta-Evaluierende müssen sich jedoch nicht etwa von der evaluierten Maßnahme oder dem Evaluationsteam fernhalten, um ihre Objektivität zu bewahren. Sie können den Kontakt mit verschiedenen am Evaluandum Beteiligten suchen und sich mit ihnen beispielsweise über den Evaluationsprozess austauschen. Dies ist gerade bei Meta-Evaluationen, die zeitgleich mit oder kurz nach der Evaluation stattfinden, in der Regel gut realisierbar. Wenn die betreffende Evaluationsstudie hingegen bereits länger zurückliegt, sind die ehemals daran beteiligten Personen häufig nicht mehr verfügbar und den Meta-Evaluatoren stehen dann lediglich Berichte und sonstige Unterlagen als Informationsquellen zur Verfügung. Die genaue Dokumentation von Evaluationsprojekten ist für Meta-Evaluationen generell wichtig (vgl.

Abschnitt 2.2.5.1) und in diesem speziellen Fall besonders relevant (ebenda).

Der spezielle Nutzen von internen Meta-Evaluationen besteht unter anderem darin, dass sie die Kommunikation zwischen den Evaluierenden und den Nutzern der Evaluation verbessern, verschiedene Perspektiven aufdecken und die Glaubwürdigkeit der Evaluatoren erhöhen. Hinzu kommt, dass die Arbeit externer Meta-Evaluatoren durch eine fundierte interne Meta-Evaluation erleichtert wird. Der zentrale Vorteil externer Meta-Evaluationen wird darin gesehen, dass externe Experten eine neue Perspektive sowie eine andere Denkweise über die Evaluation und ihren Nutzen einbringen (ebenda).

Meta-Evaluationen dienen dem Ziel, die Evaluationspraxis zu verbessern. Denn da Evaluationsprojekte jedes Jahr viele finanzielle und personelle Ressourcen binden, ist es wichtig, ihre Qualität sicherzustellen, damit diese Ressourcen bestmöglich eingesetzt werden und die Ergebnisse von Evaluationen tragfähig und nutzbar sind. Immerhin können aus mangelhaft durchgeführten Evaluationen fehlerbehaftete Entscheidungen hervorgehen (DeGEval, 2008b). In der Tat finden sich immer wieder Nachweise für Evaluationsstudien mangelnder Qualität, wie beispielsweise durch Meta-Evaluationen im Bereich der Entwicklungszusammenarbeit gezeigt werden konnte (Caspari, 2011). Zudem können auch Evaluierende sich im Nachhinein mit teilweise unberechtigten Vorwürfen konfrontiert sehen (DeGEval, 2008b). Aus diesen Gründen wird empfohlen, dass Evaluatoren ihre eigene Arbeit überprüfen (lassen), und zwar bei allen Arten von Evaluationen, also z. B. Programmevaluationen, Personalevaluationen oder Evaluationen von Produkten. Denn Meta-Evaluationen bieten Lernpotential für Evaluatoren und andere Beteiligte (Yarbrough et al., 2011) und können die Glaubwürdigkeit von Evaluationen, aber auch von Evaluatoren als Berufsgruppe erhöhen (DeGEval, 2008b). Ein weiterer Nutzen von Meta-Evaluationen kann darin gesehen werden, dass sie zur Professionalisierung von Evaluation beitragen können (Stufflebeam, 1981, 2001b; Wottawa & Thierau, 1998).

Der Wert von Meta-Evaluationen ist sowohl für die Forschungsgemeinde als auch für die Nutzer von Evaluationen bzw. ihren Ergebnissen als sehr hoch einzuschätzen. Laut Stufflebeam (2001b) sind Meta-Evaluationen für den Evaluationsbereich so wichtig wie Rechnungsprüfungen für die Buchhaltung. Jedoch sind sie auch mit Kosten

verbunden, sei es der monetäre Aufwand, um externe Meta-Evaluatoren zu finanzieren, oder der zeitliche, den auch eine knapp gehaltene, interne Meta-Evaluation erfordert. Nicht jede Evaluation muss aufwendig meta-evaluiert werden, wie Stufflebeam gerade im Hinblick auf kleinere Evaluationsprojekte mit begrenztem Wirkungskreis postuliert. Generell kann der Aufwand bei Meta-Evaluationen stark variieren. Bei der Entscheidung, ob und in welchem Ausmaß eine Meta-Evaluation realisiert wird, sollte jedoch immer beachtet werden, dass die Kosten in aller Regel den Nutzen nicht übersteigen, oder anders ausgedrückt: Meta-Evaluationen lohnen sich. Dennoch ist bisher ein Mangel an Meta-Evaluationen zu verzeichnen (Hanssen et al., 2008). Ebenso wurde bislang die Auswahl und Anwendung von Qualitätsstandards im Rahmen von Meta-Evaluationen kaum diskutiert (Cooksy & Caracelli, 2005). Forscher aus diesem Bereich fordern folglich, dass mehr Meta-Evaluationen realisiert und berichtet werden (Stufflebeam, 2001b; Widmer, 1996).

Wie aber werden Meta-Evaluationen eigentlich durchgeführt? Noch vor 30 Jahren gab es hierzu keine klaren Vorgaben und keine einheitliche Methodik (Friede, 1987). Auch heutzutage stellt Meta-Evaluation noch ein relativ unerschlossenes Forschungsgebiet dar (vgl. Abschnitt 2.2.5), wenngleich in den letzten Jahrzehnten einige Fortschritte zu verzeichnen waren. Einen Meilenstein stellte die Entwicklung von Evaluationsstandards dar, welche zur Bewertung von Evaluationsstudien herangezogen werden können. Diese stellen im Grunde eine Weiterentwicklung der für quantitative Studien gültigen Gütekriterien (Objektivität, Reliabilität sowie interne und externe Validität) dar, welche an den Evaluationskontext angepasst wurden. Da die detaillierte Beschreibung der Entwicklung dieser frühen Normen hin zu den heute gängigen Standards den Rahmen dieses Abschnitts übersteigen würde, sei der interessierte Leser auf die entsprechende Literatur verwiesen (z. B. Bührlen-Armstrong, 1998; Cook & Gruder, 1978; Greene, 1992; Schwandt & Halpern, 1988).

Die heutigen Evaluationsstandards nehmen teilweise explizit Bezug auf Meta-Evaluationen und widmen ihnen eigene Einzelstandards (DeGEval, 2008b; Yarbrough et al., 2011). In den deutschen Standards für Evaluation beispielsweise wird im Standard G9 („Meta-Evaluation“) darauf hingewiesen, dass „der Evaluationsgegenstand, die wesentlichen Zwecke, Schritte, Methoden, Daten und Ergebnisse eines Evaluations-

projektes vom Evaluationsteam dokumentiert und archiviert werden“ (DeGEval, 2008b, S. 38) sollen, um Meta-Evaluationen zu ermöglichen. Damit wird eine wichtige Voraussetzung für die Realisierung von Meta-Evaluationen benannt. Denn damit Meta-Evaluatoren eine Bewertung der Evaluation vornehmen können, müssen sie sich zunächst möglichst detailliert mit deren Abläufen und Inhalten vertraut machen. Wittmann (1985) weist darauf hin, dass hierbei die Bereitschaft der Evaluatoren zur Archivierung und Herausgabe ihrer Daten einen entscheidenden Faktor darstellt und auch datenschutzrechtliche Bedenken eine Rolle spielen können. Andererseits könnte die Verpflichtung zur Datensammlung und Dokumentation bereits für sich genommen eine höhere Qualität der Forschungsprozesse bewirken.

Was das konkrete Vorgehen bei einer Meta-Evaluation betrifft, so gibt es bislang relativ wenige Publikationen, die sich mit diesem Thema beschäftigen. Im folgenden Abschnitt werden diese im Kontext des Forschungsstandes in Bezug auf Meta-Evaluationen und Evaluationsstandards vorgestellt.

2.2.5 Forschungsstand bezüglich Meta-Evaluationen und Evaluationsstandards

In diesem Abschnitt wird der aktuelle Stand der Forschung hinsichtlich Meta-Evaluationen und Evaluationsstandards wiedergegeben. Beide Themen sind miteinander verbunden, stellen doch Evaluationsstandards eine mögliche Bewertungsrichtlinie für Meta-Evaluatoren dar. Daher können auch die Forschungsergebnisse zu diesen beiden Bereichen nicht ganz trennscharf, d. h. ohne Überschneidungen berichtet werden. Der Abschnitt beginnt mit einer Darstellung von Publikationen zum Vorgehen und der Methodik bei Meta-Evaluationen. Danach werden Befunde speziell zu den amerikanischen sowie den deutschen Evaluationsstandards referiert. Abschließend wird der Forschungsstand zusammengefasst und auf die Fragestellungen der vorliegenden Arbeit übergeleitet.

Bereits einleitend sei darauf hingewiesen, dass die Literatur zu den Themen „Meta-Evaluation“ und „Evaluationsstandards“ wenig umfangreich ist (vgl. Bührlen-Armstrong, 1998). Der Großteil der Studien und Aufsätze stammt dabei aus dem anglo-amerikanischen Raum, deutsche Publikationen sind vergleichsweise rar gesät. Auch in

deutschsprachigen Lehrbüchern, die sich mit Evaluation beschäftigen, wurde und wird die Thematik der Meta-Evaluation wenig ausführlich behandelt (Bortz & Döring, 2006; Döring, 2009; Wittmann, 1985; Wottawa & Thierau, 1998). Dies steht im Einklang mit der Tatsache, dass es sich um ein noch junges Forschungsgebiet handelt und unterstreicht zugleich den Bedarf an weiterer Forschungsaktivität in diesem Bereich.

2.2.5.1 Methodik bei Meta-Evaluationen: Generelles Vorgehen

Zunächst kann das Vorgehen im Rahmen einer Meta-Evaluation grob danach unterschieden werden, ob dabei (bereits vorhandene oder neu erhobene) Daten analysiert werden oder nicht (Cook & Gruder, 1978; Wittmann, 1985). Wenn keine Datenanalyse erfolgt, besteht das Vorgehen in der Regel in der Untersuchung von Dokumenten (z. B. Projektberichten), weshalb Meta-Evaluatoren besonders auf die Zugänglichkeit und Qualität dieser Dokumente angewiesen sind (Yarbrough et al., 2011). Zudem werden teilweise Interviews mit an der Evaluation beteiligten Personen geführt, um die Wissensbasis der Meta-Evaluatoren zu erweitern (z. B. Widmer, 1996). Die Erhebung und Auswertung von Daten bzw. die Re-Analyse bereits vorliegender Daten kann je nach Fragestellung erforderlich sein, wenn beispielsweise der langfristige Nutzen einer Evaluation festgestellt werden soll (Yarbrough et al., 2011).

Wie bereits erwähnt, stellt die Qualität der Dokumentation von Evaluationsprojekten eine wichtige Grundlage für Beurteilungen durch Meta-Evaluatoren dar. So ist es wichtig, dass der Evaluationsprozess detailliert festgehalten und das Datenmaterial archiviert wird (Wittmann, 1985; Yarbrough et al., 2011). Generell sollten Evaluationen auf eine Weise berichtet bzw. dokumentiert werden, die spätere Meta-Evaluationen ermöglicht. Ein solches Vorgehen wäre auch dann nützlich, wenn eine ähnliche Maßnahme bzw. Evaluation entworfen werden soll und die Evaluationsberichte eine Orientierungshilfe bieten können. Entsprechende Empfehlungen sind in den deutschen (DeGEval, 2008b) und amerikanischen (Yarbrough et al., 2011) Evaluationsstandards enthalten und werden auch von weiteren Autoren ausgesprochen (z. B. Widmer, 1996). Jedoch gibt es etliche Belege dafür, dass die geforderte detaillierte Dokumentation von Evaluationen (noch) nicht ausreichend realisiert wird (z. B. Caspari, 2010; Scott-

Little, Hamann & Jurs, 2002; Yarbrough et al., 2011). Wünschenswert sind zudem auch Publikationen, die neben den Verfahren und Ergebnissen von Evaluationen auch Probleme thematisieren (DeGEval, 2008b). Denn die Erfahrungen, die in Evaluationsstudien gemacht werden, könnten genutzt werden, um andere Evaluationen zu optimieren. Dieses Lernpotential wird jedoch bislang fast nie ausgeschöpft (Spiel et al., 2012).

Daniel Stufflebeam hat eine Methodik entworfen, die das Vorgehen bei einer Meta-Evaluation in 11 Aufgaben unterteilt (Stufflebeam, 2001b). Diese können Tabelle 1 entnommen werden.

Tabelle 1: Aufgaben im Verlauf einer Meta-Evaluation

Aufgabe (Nummer)	Beschreibung der Aufgabe
1	Stakeholder der ME ermitteln und Kontakt herstellen
2	MEs-Team mit einem oder mehreren qualifizierten Evaluatoren besetzen
3	Fragestellungen der ME definieren
4	Auf Standards, Prinzipien und/oder Kriterien einigen, anhand derer die Evaluation beurteilt werden soll
5	Vertrag entwickeln, der die ME regeln soll
6	Geeignete verfügbare Informationen beschaffen und einsehen
7	Bei Bedarf neue Informationen erheben (z. B. Interviews oder Beobachtungen)
8	Qualitative und quantitative Daten analysieren
9	Einhaltung geeigneter Standards, Prinzipien und/oder Kriterien durch die Evaluation beurteilen
10	Ergebnisse der ME durch Berichte, mündliche Präsentationen etc. vermitteln
11	Falls notwendig und durchführbar, den Klienten und andere Stakeholder bei der Interpretation und Anwendung der Ergebnisse unterstützen

Anmerkungen. ME = Meta-Evaluation
(nach Stufflebeam, 2001b, S. 191)

Die einzelnen Aufgaben werden hier aus Platzgründen nicht näher besprochen. Stufflebeams Methodik hat den Vorteil, dass sie den

meta-evaluativen Vorgang chronologisch strukturiert und daher eine anschauliche und nachvollziehbare Orientierung bietet. Der Autor weist darauf hin, dass nicht alle der genannten Aufgaben bei jeder Meta-Evaluation vorkommen müssen und vermutlich bei einigen Meta-Evaluationen noch weitere Aspekte zu ergänzen sind (ebenda).

2.2.5.2 Methodik bei Meta-Evaluationen: Anwendung von Evaluationsstandards

Wie wird nun aber der zentrale Aspekt der Meta-Evaluation, die Bewertung der Evaluationsstudie, umgesetzt? Es wurde bereits erwähnt, dass (Evaluations-)Standards zur Bewertung herangezogen werden können. Stufflebeams Modell (vgl. Tabelle 1) zufolge sind zuerst geeignete Standards oder Prinzipien auszuwählen (Aufgabe 4), um nachfolgend deren Einhaltung zu prüfen (Aufgabe 9).

Welche Standards geeignet sind, hängt von der jeweiligen Situation ab (Stufflebeam, 2001b). Es können z. B. die unter Abschnitt 2.2.2 besprochenen Evaluationsrichtlinien verwendet werden, die teilweise durch andere Standards ergänzt werden können, sofern dies als notwendig und sinnvoll erachtet wird. So wird in den amerikanischen Standards des JCSEE empfohlen, Meta-Evaluationen anhand der JCSEE-Standards *und* anderer geeigneter Richtlinien durchzuführen (Yarbrough et al., 2011). Eine analoge Empfehlung findet sich auch für die DeGEval-Standards (DeGEval, 2002). Unter den wenigen veröffentlichten Artikeln, die über Meta-Evaluationen berichten, finden sich demnach Studien, in denen nur die amerikanischen Standards des JCSEE (Finn, Jr., Stevens, Stufflebeam & Walberg, 1997; Lynch et al., 2003; Ruhe & Boudreau, 2013; Scott-Little et al., 2002) oder nur die deutschen Richtlinien der DeGEval (Schendera, Janz, Klauß & Lamers, 2003) verwendet wurden neben Untersuchungen, in denen diese durch weitere Standards ergänzt wurden (Caspari, 2010, 2011; Hanssen et al., 2008). Bei zwei deutschsprachigen Dissertationen wurden zudem eigene Kriterienraster entwickelt (basierend unter anderem auf den JCSEE-Standards), Evaluationsstandards aus dem deutschen Sprachraum lagen zu dieser Zeit noch nicht vor (Bührlen-Armstrong, 1998; Widmer, 1996). Es wird empfohlen, die im Rahmen der Meta-Evaluation eingesetzten Bewertungskriterien vorab mit den Beteiligten zu besprechen und hier einen Konsens zu finden (Widmer, 1996).

Auf die Frage, wie die ausgewählten Standards zur Bewertung von Evaluationen eingesetzt werden können, liefert die vorhandene Literatur nur unzureichende Antworten. In den Evaluationsstandards selbst sind lediglich allgemeine Hinweise enthalten, wie beispielsweise, dass die Anwendung der Standards nicht schematisch erfolgen sollte und auch Widersprüche zwischen Standards möglich sind (vgl. Abschnitt 0). Das erscheint für die Anwendung der Richtlinien bei der Planung und Durchführung von Evaluationen plausibel, da hier der jeweilige Kontext eine wichtige Rolle spielt und ein einheitliches Vorgehen im Sinne der starren Anwendung aller Standards als wenig sinnvoll zu erachten ist. Für Meta-Evaluatoren fehlt jedoch eine Orientierung, wie die Bewertung erfolgen soll und was dabei zu beachten ist.

Auch der Einbezug weiterer Literatur gibt hierzu wenig Aufschluss. Die hilfreichsten Hinweise stammen aus dem anglo-amerikanischen Raum und beziehen sich auf die Standards des JCSEE. Auch in diesem Punkt kann Stufflebeam als Vorreiter bezeichnet werden. Er hat für die Standards des JCSEE eine Methodik erarbeitet, nach welcher jeder Einzelstandard im Hinblick darauf untersucht wird, ob er eingehalten wurde oder nicht. Dazu wurden für jeden Einzelstandard sechs bis zehn Unterpunkte erarbeitet, anhand derer einzuschätzen ist, ob der Standard als erfüllt, teilweise erfüllt oder nicht erfüllt angesehen wird. Auch die vier Oberkategorien (*Utility*, *Feasibility*, *Propriety* und *Accuracy*) können dahingehend beurteilt werden, inwiefern diese eingehalten wurden, je nachdem, wie die Einstufung der zugehörigen Einzelstandards ausfällt (zum genaueren Vorgehen vgl. Stufflebeam, 2001b).

Kritisch anzumerken ist allerdings, dass die Entstehung der Unterpunkte nicht erläutert wird und daher etwas willkürlich anmutet. Zudem bezieht sich das Verfahren nicht auf die aktuelle Version der JCSEE-Standards von 2011. Die Methodik von Stufflebeam hat den Vorteil, dass sie eine Struktur für die Bewertung einer Evaluation anhand der Evaluationsstandards bietet. Immerhin füllen die Standards des JCSEE ein über 300 Seiten umfassendes Buch, sodass Meta-Evaluatoren zwar einerseits mit vielen nützlichen Hinweisen versorgt werden, andererseits aber womöglich überfordert sind, wenn es darum geht, aus dieser Vielzahl an Informationen die Bewertung einer Evaluation abzuleiten. Ein großer Nutzen von Stufflebeams Vorgehen besteht darin, diese Komplexität zu verringern.

Auch die Checklisten, die ebenfalls von der Forschergruppe um Stufflebeam entwickelt wurden, dienen diesem Zweck. Neben weiteren Zielsetzungen, die bereits in Abschnitt 2.2.2 dargelegt wurden, können solche Checklisten ein nützliches Instrument für die Durchführung von Meta-Evaluationen darstellen (Stufflebeam, 2001b). Als Beispiel sei die Checkliste für die amerikanischen Standards des JCSEE beschrieben (Stufflebeam, 1999a, 2000). Auch hier werden für jeden Einzelstandard 10 Unterpunkte danach beurteilt, ob sie eingehalten wurden oder nicht. Die Bewertung des Einzelstandards erfolgt über die Zuteilung eines Punktwertes, der von der Anzahl eingehaltener Unterpunkte abhängt. So wird die Bewertung *ausgezeichnet* bei neun oder zehn erfüllten Unterpunkten vergeben, das Urteil *sehr gut* bei sieben oder acht, *gut* bei fünf oder sechs, *ausreichend* bei drei oder vier, und *mangelhaft* bei null bis zwei Punkten. Für die Einschätzung der Oberkategorien (*Utility* etc.) wird die Anzahl der *ausgezeichnet*-Urteile addiert und mit dem Wert vier multipliziert, selbiges für die übrigen Urteile, wobei *sehr gut* drei Punkte ergibt, *gut* zwei und so fort. Die multiplizierten Urteile werden aufsummiert und ergeben einen Wert, der anhand einer Skala wiederum in ein Urteil für die Oberkategorie umgewandelt wird (z. B. bei 26 erreichten von 28 möglichen Punkten wird die Oberkategorie als *ausgezeichnet* eingestuft).

Problematisch erscheint der Einsatz solcher Checklisten im Rahmen von Meta-Evaluationen dann, wenn sie allzu strikt befolgt und nicht an die Erfordernisse der Situation angepasst werden. Denn immerhin wird in den gängigen Evaluationsstandards selbst darauf hingewiesen, dass nicht bei jeder Evaluation alle Standards immer einhaltbar sind (DeGEval, 2008b; Yarbrough et al., 2011; vgl. auch DeGEval, 2002). Genau dies wäre aber die Voraussetzung bei der Berechnung von Summenwerten wie im obigen Beispiel. Die Checklisten sollten also mit Bedacht eingesetzt und ggf. adaptiert werden, indem Standards, die in einem gegebenen Kontext nicht erfüllt werden *können*, bei der Berechnung entsprechend berücksichtigt werden. Zu beachten ist auch, dass Stufflebeam keine Gewichtung vornimmt und alle Einzelstandards gleichgewichtig in die Bewertung einfließen. Allerdings enthalten auch die Evaluationsstandards keine Hinweise zu einer möglichen Gewichtung, sodass eine solche im Einzelfall von den Meta-Evaluatoren selbst ausgearbeitet werden müsste. Im Übrigen sollten die Checklisten nicht als einziges Instrument verwendet werden, sondern vielmehr durch Informationen z. B. aus den

Handbüchern zu Evaluationsstandards ergänzt werden. Denn sie können die Arbeit der Meta-Evaluierenden zwar erleichtern, ihnen aber die intensive Auseinandersetzung mit den Evaluationsstandards nicht abnehmen. Ein besonderer Vorteil solcher Checklisten besteht darin, dass sie alle zu berücksichtigenden Aspekte abdecken und somit das Vergessen einzelner Aspekte verhindern helfen (Yarbrough et al., 2011).

Auffällig ist, dass die hier dargestellte Methodik (einschließlich der Checklisten) nur für die amerikanischen Evaluationsstandards existiert. Ein Pendant für die deutschen DeGEval-Standards wurde bislang nicht erarbeitet.

Der Einsatz von Evaluationsstandards oder anderen Kriterien bei der Bewertung von Evaluationen fördert die Systematik des Vorgehens. Jedoch bleibt die Einschätzung, ob bzw. inwieweit ein Standard erfüllt wurde, ein subjektiver Vorgang (Widmer, 1996). In einigen der Meta-Evaluationen wurde daher diese Einschätzung durch zwei Personen unabhängig voneinander vorgenommen, um nachfolgend die Inter-rater-Reliabilität berechnen zu können und somit eine gewisse Objektivität der Urteile zu gewährleisten (Bührlen-Armstrong, 1998; Scott-Little et al., 2002). Leider wird beim Großteil der gefundenen Meta-Evaluationen die angewendete Methodik derart knapp beschrieben, dass sich daraus nicht erkennen lässt, ob die Urteile durch einen zweiten Rater abgesichert wurden (Caspari, 2010, 2011; Finn et al., 1997; Lynch et al., 2003; Ruhe & Boudreau, 2013; Schendera et al., 2003). Daher sind auch keine fundierten Aussagen zum Vorteil dieser Methodik gegenüber der Bewertung durch nur eine Person möglich.

Als Nachteil externer Evaluationsbewertungen ist zu benennen, dass den Meta-Evaluatoren oftmals nicht genügend Informationen zur Verfügung stehen, um für alle Standards verlässliche Aussagen hinsichtlich ihrer Einhaltung treffen zu können (z. B. Scott-Little et al., 2002; Widmer, 1996). Die mangelhafte Beschreibung der in Meta-Evaluationen angewandten Methodik bringt allerdings mit sich, dass meist nicht klar ersichtlich wird, ob es sich um eine interne oder um eine externe Meta-Evaluation handelt. Ebenso fehlen häufig Anmerkungen der Autoren dazu, welche Erfahrungen sie mit dem jeweiligen Vorgehen gemacht haben sowie welche Empfehlungen sie daraus ableiten. Gerade in Anbetracht der Tatsache, dass Meta-Evaluation ein

junges Forschungsfeld darstellt, würden solche Praxisberichte und Anregungen einen wichtigen Beitrag zu dessen Weiterentwicklung leisten.

Abschließend wird die Meta-Evaluation von Thomas Widmer (1996) etwas näher beschrieben, da die in der vorliegenden Studie eingesetzte Methode sich daran orientiert (Abschnitt 4.2.1). Widmer bewertete 10 schweizerische Evaluationen, die in unterschiedlichen Politikfeldern durchgeführt worden waren, anhand eines Kriterienrasters, der stark an die amerikanischen Evaluationsstandards von 1981 angelehnt war und 30 Bewertungskriterien umfasste. Die Studie war ursprünglich in Form einer Dokumentenanalyse geplant gewesen, wobei für zwei der zehn Evaluationen eine vertiefte Untersuchung durch Interviews mit Beteiligten vorgesehen war. Da die zur Verfügung stehenden Dokumente jedoch für eine verlässliche Bewertung nicht ausreichten, wurden schließlich mit Beteiligten aller zehn Evaluationen Interviews geführt. Die dabei erhaltenen Informationen wurden zusammen mit den Daten aus den Unterlagen den 30 Bewertungskriterien zugeordnet und Fallstudien erstellt. Um Fehleinschätzungen und Lücken bei der Beurteilung zu vermeiden und außerdem die Evaluatoren einzubeziehen, wurden diesen die Fallstudien vorgelegt und sie wurden um Rückmeldungen gebeten. Diese führten zu einer Überarbeitung der Fallstudien und nachfolgend wurde den Evaluatoren noch die Möglichkeit gegeben, ihre Sichtweise in einer Stellungnahme auszudrücken, die den Fallstudien in unveränderter Form beigefügt wurde. Um den subjektiven Bewertungsprozess nachvollziehbarer zu machen, entschied sich Widmer außerdem dafür, die einzelnen Bewertungen detailliert zu schildern und zu begründen.

Widmers Arbeit kann eine Vorreiterrolle in Bezug auf Meta-Evaluationen im deutschsprachigen Raum zugeschrieben werden. Sie erbrachte zudem zahlreiche Empfehlungen zu Evaluationen und Meta-Evaluationen, die auch zur Entwicklung der schweizerischen Evaluationsstandards beitrugen.

2.2.5.3 Befunde zu den amerikanischen Evaluationsstandards des JCSEE

Nach der Schilderung von Erkenntnissen zur Methodik bei Meta-Evaluationen werden im Folgenden weitere Informationen und Befunde

speziell zu den JCSEE-Standards zusammengetragen. Dies geschieht vor dem Hintergrund, dass zu diesen Standards mehr Forschung betrieben wurde als zu ihrem deutschen Pendant. Bei der in der vorliegenden Arbeit vorgestellten Meta-Evaluation wurden die DeGEval-Standards sowie ergänzend die amerikanischen Richtlinien verwendet. Da die deutschen Standards den amerikanischen sehr ähnlich sind, ist es wahrscheinlich, dass Befunde, die sich auf die JCSEE-Standards beziehen, auf die deutschen Standards übertragbar sind.

Eingangs sei erwähnt, dass in der aktuellen Version der amerikanischen Evaluationsstandards der neue Bereich *Evaluation Accountability* mit aufgenommen wurde (Yarbrough et al., 2011; vgl. Abschnitt 2.2.2). Dieser umfasst drei Standards, die auf die Dokumentationspflicht als Grundlage für Meta-Evaluationen, auf interne sowie auf externe Meta-Evaluationen eingehen. Die Tatsache, dass die Thematik der Meta-Evaluation im Vergleich zur vorherigen Version deutlich detaillierter behandelt wird, spricht für eine steigende Relevanz dieses Bereichs in den letzten Jahren.

Die amerikanischen Standards sind (ebenso wie die Richtlinien der DeGEval) als Maximalstandards konzipiert, die nicht bei jeder Evaluation gleichermaßen eingehalten werden können. Dies ist auch darauf zurückzuführen, dass die Standards teilweise im Widerspruch zueinander stehen (vgl. Abschnitt 0). Aber welche Standards erweisen sich in der Praxis als unvereinbar? Gibt es Standards, die häufiger eingehalten werden als andere? Die bisherige Forschung erbringt uneinheitliche Ergebnisse. So ließen Newman und Brown (1992) in ihrer Untersuchung Personen mit unterschiedlichen Erfahrungen im Evaluationsbereich einschätzen, wie häufig in der Praxis gegen die Standards des JCSEE (in der ersten Auflage von 1981) verstoßen wird. Zudem sollte die Bedeutsamkeit der Verstöße jeweils bewertet werden. Es zeigte sich, dass für die Kategorien *Accuracy* und *Propriety* die wenigsten und für *Utility* und *Feasibility* die meisten Verstöße angegeben wurden. Die Einschätzungen bezüglich der Bedeutsamkeit der Nicht-Einhaltung verhielten sich umgekehrt: So wurden Verletzungen der *Accuracy*- und der *Propriety*-Standards am gravierendsten beurteilt, Verstöße gegen die beiden anderen Gruppen hingegen als weniger bedeutsam angesehen. Die Autoren regen weitere Forschung dazu an, wie auftretende Konflikte zwischen Standards gelöst werden. In der Meta-Evaluation von Finn und Kollegen wurde die Evaluation eines

Lernprogramms in Schulen anhand der JCSEE-Standards (ebenfalls in der ersten Auflage von 1981) evaluiert (Finn et al., 1997). Diese Studie erbrachte gute Ergebnisse bezüglich der Einhaltung der *Accuracy*-, *Propriety*- und *Feasibility*-Standards, lediglich im *Utility*-Bereich waren Abstriche zu verzeichnen. Die Meta-Evaluation von Widmer (1996) bewertete zehn Evaluationsstudien anhand eines an den JCSEE-Standards von 1981 orientierten Kriterienrasters. Hier ergaben sich für die *Utility*-, *Feasibility*- und *Propriety*-Standards positive Bewertungen, während die *Accuracy*-Richtlinien für nicht ganz so gut eingehalten befunden wurden. Eine weitere Meta-Evaluation befasste sich mit Evaluationen von Programmen im Bereich der Nachmittagsbetreuung (*after-school programs*) und nutzte die JCSEE-Standards in ihrer zweiten Version von 1994 (Scott-Little et al., 2002). Die bewerteten Evaluationsstudien hatten viele Standards erfüllt, einige jedoch nicht eingehalten. Zudem war bei einigen Standards keine Beurteilung möglich, da die Evaluationsberichte keine ausreichenden Informationen enthielten. Keine Evaluation hatte alle Standards erfüllt. Insgesamt waren die Standards in den Bereichen *Accuracy* und *Utility* besser eingehalten worden als in den Bereichen *Feasibility* und *Propriety*. Da die Autoren die Einhaltung der Standards aufgrund fehlender Informationen in einigen Fällen nicht bewerten konnten, ziehen sie die Schlussfolgerung, dass zwar alle Standards wichtig sind, aber nicht für alle Zielgruppen in gleichem Maße. Dies könnte der Grund sein, warum die Evaluationsberichte viele Angaben, die für Meta-Evaluatoren wichtig wären, nicht enthalten, weil sie für die Zielgruppe des Berichts nicht von gleicher Relevanz sind. Schließlich sei die relativ aktuelle Meta-Evaluation von Ruhe und Boudreau (2013) genannt, in der ein Lernprogramm im Rahmen des Medizinstudiums anhand der JCSEE-Standards in ihrer jüngsten Fassung von 2011 bewertet wurde. Während die Autoren nicht im Detail darauf eingehen, welche Standards erfüllt wurden, wird erwähnt, dass zwischen verschiedenen Standardgruppen Widersprüche auftraten, beispielsweise zwischen den Bereichen *Propriety* und *Feasibility*.

Die beschriebenen Ergebnisse können dahingehend integriert werden, dass die *Accuracy*- und die *Propriety*-Standards überwiegend beachtet und eingehalten wurden, während im *Utility*-Bereich und bei den *Feasibility*-Standards mehr Verstöße zu verzeichnen waren. Diese Ergebnisse passen zu dem Befund von Newman und Brown (1992), wonach eine Verletzung der *Accuracy*- und der *Propriety*-Standards am

gravierendsten beurteilt wurde. Dass die *Accuracy*-Standards mit am besten befolgt wurden, könnte damit erklärt werden, dass diese am ehesten mit den in der Wissenschaft etablierten Richtlinien übereinstimmen (z. B. die Forderung nach reliablen und validen Daten, vgl. Abschnitt 2.2.3.4) und daher von Forschern intuitiv eingehalten werden, während die übrigen Bereiche eher im speziellen Kontext der Evaluation Anwendung finden (vgl. Schiffler & Hübner, 2000).

Ein weiterer Aspekt, dem leider nur in zwei Studien Rechnung getragen wurde, ist die Eignung der Evaluationsstandards für die Durchführung von Meta-Evaluationen. So heben Lynch und Kollegen hervor, dass sie die JCSEE-Standards als gutes Instrument für ihre Meta-Evaluation empfanden, welches umfassend ist und alle wichtigen Aspekte abdeckt (Lynch et al., 2003). Bei dieser Untersuchung wurde die Evaluation eines Trainingsprogrammes im Gesundheitsbereich anhand der 1994er-Standards (2. Auflage) bewertet. Den Autoren zufolge halfen die Standards dabei, Stärken und Schwächen der Evaluation zu erörtern und sie nachfolgend zu verbessern. Insbesondere wurden durch die Verwendung der Standards Aspekte betont, die sonst unbeachtet geblieben wären (z. B. die Identifizierung von Werten). Lynch et al. weisen zudem darauf hin, dass die meisten Evaluationsstudien die *Accuracy*-Standards befolgen, was sich mit den oben genannten Befunden deckt. Die Autoren halten jedoch die anderen Standardgruppen ebenfalls für wichtig. Auch in der bereits genannten Meta-Evaluation von Ruhe und Boudreau (2013) erwiesen sich die Standards als hilfreich und umfassend. Die Autoren hatten selbst auch die zugrunde liegende Evaluation durchgeführt, die sie nun retrospektiv bewerteten. Im Rahmen der Meta-Evaluation wurden einige Herausforderungen offenbart, die ihnen während der Evaluation verborgen gewesen waren. Die Autoren halten fest, dass sie diese Probleme bereits im Verlauf der Evaluation hätten erkennen und lösen können, wenn sie die Evaluationsstandards von vornherein beachtet hätten. Folglich empfehlen die Autoren, bei künftigen Evaluationen die Standards in allen Phasen anzuwenden. Auch Stufflebeam (1999b) rät dazu, dass die Standards vom Beginn einer Evaluation an eingesetzt werden sollten.

Die amerikanischen Standards sind nicht unumstritten, sondern werden teils auch kritisch gesehen. So hält Raven (1984) die JCSEE-Standards in ihrer ersten Auflage von 1981 für wenig hilfreich, weil sie

zu weit vom Alltag der Evaluatoren entfernt sind und unrealistische Erwartungen an diese stellen. Und auch Roth (2012) bemängelt die Realitätsferne der amerikanischen Standards, die einerseits selbstverständliche und andererseits unrealistische Anforderungen enthalten. Diesen kritischen Stimmen stehen die eben angeführten positiven Erfahrungen einzelner Autoren mit den Standards bei deren Anwendung im Rahmen von Meta-Evaluationen gegenüber. Es scheint daher angebracht, diesbezüglich weitere Forschungsaktivitäten zu fordern, um die Angemessenheit der Standards in der Praxis zu prüfen und gegebenenfalls Ansatzpunkte für deren Überarbeitung in Richtung einer größeren Realitätsnähe herauszufinden. Tatsächlich bitten auch die Autoren der JCSEE-Standards selbst darum, dass die Anwender der Standards ihre Erfahrungen teilen und die Angemessenheit und Anwendbarkeit der Standards bewerten, damit diese kontinuierlich verbessert werden können (Yarbrough et al., 2011).

2.2.5.4 Befunde zu den deutschen Standards für Evaluation

Zu den Evaluationsstandards der DeGEval existiert noch weniger Forschungsliteratur als zu ihrem amerikanischen Pendant. Das mag darin begründet liegen, dass die JCSEE-Standards gut 20 Jahre vor den DeGEval-Standards erschienen sind, ist aber sicher ebenso darauf zurückzuführen, dass der Bereich der Meta-Evaluation in Deutschland noch nicht die gleiche Aufmerksamkeit bekommt wie im anglo-amerikanischen Sprachraum. Im Einklang damit steht die Tatsache, dass in der aktuellen Version der JCSEE-Standards die Thematik der Meta-Evaluation mehr Beachtung findet (vgl. Abschnitt 2.2.5.3), während diesem Aspekt in den Richtlinien der DeGEval bisher nur ein Einzelstandard gewidmet ist.

Zunächst seien die Befunde einer Befragung dargestellt, die die DeGEval 2004 durchführte, um Informationen für den Revisionsprozess ihrer Evaluationsstandards zu erhalten (DeGEval, 2004a, 2004b). Dabei wurden mit Evaluationen befasste Personen online befragt, darunter Evaluatoren, Wissenschaftler mit Interesse an Evaluation, Berater und auch Auftraggebende von Evaluationen. Von denjenigen Befragten, die die DeGEval-Standards zur Verfügung hatten, hatten drei Viertel diese anderen Personen empfohlen, was dafür spricht, dass die Standards mehrheitlich als nützlich betrachtet wurden. Die Standards waren von

vielen Befragten als allgemeine Orientierungsgrundlage sowie als phasenübergreifendes Hilfsmittel bei Evaluationen genutzt worden. Ebenso hatten sie Anwendung im Rahmen der Qualitätssicherung gefunden, wobei sie als Grundlage von Meta-Evaluationen oder als Hilfsmittel zur Selbstreflexion fungierten. Des Weiteren waren die Standards als Legitimationsgrundlage oder für Lehrzwecke verwendet worden.

Da die Befragung auf eine mögliche Revision der Standards abzielte, wurden die Teilnehmenden um Rückmeldung zu den Standards selbst gebeten. Es zeigte sich, dass neben formalen Veränderungen wie z. B. einer anderen Publikationsform auch inhaltliche Verbesserungen gewünscht wurden, darunter die Aufnahme von Fallbeispielen, häufigen Fehlern, Lösungen bei Problemen bezüglich der Anwendung der Standards sowie Hinweise und Ratschläge zur Interaktion zwischen Evaluierenden und Auftraggebenden. Diese Vorschläge wiesen in Richtung einer deutlichen Erweiterung der DeGEval-Standards hin zu einem ausführlichen Handbuch. Die amerikanischen Standards des JCSEE haben die Form eines solchen ausführlichen Handbuchs mit Fallbeispielen, häufigen Fehlern etc. Da viele der Befragten die JCSEE-Standards kannten, steht zu vermuten, dass sie diese Form auch für die deutschen Standards bevorzugen würden und daher die genannten Anregungen für eine Revision machten. Allerdings sahen etliche der Befragten auch keinen Bedarf für eine Revision oder wünschten sich nur Änderungen von geringerem Umfang (ebenda).

Weiterhin wurden in der Befragung teils widersprüchliche Forderungen geäußert, die die Autoren als offene Fragen („issues“) zusammenfassten. So wurde von einigen Befragten die Formulierung der Standards als Maximalstandards anerkannt und für sinnvoll erachtet, während andere Teilnehmer sich Minimalstandards wünschten. Ebenso wurde teils eine knappere, teils jedoch auch eine detailliertere Fassung als Wunsch genannt. Während zudem einige der Befragten die Allgemeinheit der Standardformulierungen schätzten, äußerten vergleichsweise viele Teilnehmende einen Bedarf nach feldspezifischen Standards, die an die jeweiligen Anwendungsbereiche angepasst sind. Neben positiven Bemerkungen gab es auch negative Kommentare, darunter Kritik an der Praxisferne der Standards und damit an einem Punkt, der mitunter auch an den JCSEE-Standards bemängelt wird (vgl. Abschnitt 2.2.5.3). Schließlich wurden inhaltliche Rückmeldungen zu einzelnen Standards gegeben, wobei zu beachten ist, dass diese

lediglich von einzelnen Personen stammen und deren Erfahrung und Meinung widerspiegeln. So wurde z. B. verlangt, dass auch die nicht-methodologischen Standards bekannter gemacht werden, da in bisherigen Evaluationsberichten überwiegend die methodologischen Genauigkeitsstandards erfüllt seien. Diese Forderung stimmt mit den im vorigen Abschnitt geschilderten Erkenntnissen zu den amerikanischen Standards überein. Hinsichtlich des Standards N4 (Auswahl und Umfang der Informationen: Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellung zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen) lautete eine Rückmeldung, dass der Umgang mit diesem Standard schwierig sei, da man den Auftraggeber klar berücksichtige, bei anderen Adressaten aber Schwierigkeiten sehe. Zusätzlich zu diesen inhaltlichen Anmerkungen wurden Änderungen der Formulierungen einzelner Standards vorgeschlagen, die an dieser Stelle nicht näher ausgeführt werden (ebenda).

Ergänzend sei erwähnt, dass der Revisionsprozess, dem die geschilderte Befragung der DeGEval gewidmet war, aktuell noch andauert. In der Neuauflage der DeGEval-Standards im Jahr 2008 waren lediglich geringfügige Änderungen vorgenommen worden[5]. Die DeGEval befürwortet in diesem Zusammenhang Meta-Evaluationen, in denen die Standards zur Anwendung kommen, da diese einen Beitrag zur fachlichen Qualifizierung des Revisionsprozesses leisten können (DeGEval, 2002).

In einer weiteren Untersuchung wurden Tiefeninterviews mit im Evaluationsbereich tätigen Akteuren geführt, die überwiegend innerhalb der DeGEval aktiv waren (Brandt, 2009). Diese sahen den Nutzen der Standards bislang in erster Linie darin, eine Orientierung zum Thema Evaluation zu bieten sowie die Planung und Durchführung von Evaluationsstudien zu unterstützen. Teilweise wurde auch die Funktion der Standards als Argumentationsgrundlage benannt. Nach der Relevanz der deutschen Evaluationsstandards gefragt, schätzten die meisten Befragten diese als hoch ein und sahen die Standards als notwendig an, um die Transparenz der Evaluationspraxis zu erhöhen. Seltener und nur von Nicht-Mitgliedern der DeGEval wurde die Relevanz als gering

[5] Diese Information entstammt einer elektronischen Korrespondenz mit der Geschäftsstelle der DeGEval im Mai 2015.

bewertet. Dies geschah mit der Begründung, die Standards seien nicht spezifisch genug, um konkretes Handeln in den verschiedenen Praxisfeldern anzuleiten. Anderen, verbindlicheren Richtlinien wurde eine höhere praktische Relevanz zugeschrieben. Der gegenwärtige praktische Stellenwert und die Verbreitung der Standards wurden insgesamt und insbesondere auf Auftraggeberseite als gering angesehen. Als mögliche Erklärung wurde der Umfang der Standards genannt, die eine intensive inhaltliche Einarbeitung erfordern würden und nicht im Sinne eines Checklistenverfahrens schnell und einfach einsetzbar seien. Eine Verbreitung wie in der Schweiz, wo der Einbezug der SEVAL-Standards bei Evaluationen gängig sei, sei in Deutschland wünschenswert, aber noch weit entfernt. In diesem Zusammenhang kritisierten einige der Befragten auch eine fehlende inhaltliche Diskussion der deutschen Standards, die, in Verbindung mit einer besseren Zugänglichkeit und Publikation von Evaluationsberichten, einen Beitrag zur Qualitätssicherung leisten könne. Als weitere Maßnahmen im Sinne der Qualitätssicherung wurde einerseits die Durchführung von mehr Meta-Evaluationen (basierend auf den DeGEval-Standards oder vergleichbaren Richtlinien) gefordert und andererseits eine Sensibilisierung bzw. Qualifizierung der Auftraggebenden von Evaluationen gewünscht.

Die bisher skizzierten Forschungsarbeiten zu den DeGEval-Standards bilden Meinungen von Experten aus dem Evaluationsbereich ab und liefern damit wertvolle Informationen. Jedoch fehlen bislang weitgehend Berichte über Studien, in denen die Standards tatsächlich eingesetzt wurden und diesbezügliche Erfahrungen und mögliche Herausforderungen behandelt werden. Auch zur Eignung der deutschen Standards für die Durchführung von Evaluationen oder Bewertungen im Rahmen von Meta-Evaluationen finden sich so gut wie keine Befunde.

Die einzige gefundene Untersuchung, in der die Umsetzung der 25 DeGEval-Standards bei einem konkreten Evaluationsprojekt beschrieben wird, ist die Studie von Schendera und Kollegen (Schendera et al., 2003). Dort wird jedoch weder auf die Methodik der Bewertung noch auf die Eignung der Standards oder gefundene Widersprüche eingegangen. Auffällig ist vielmehr, dass den Autoren zufolge alle Standards eingehalten wurden bzw. künftig noch werden, sodass fraglich erscheint, wie kritisch ihre Einschätzung erfolgte, zumal es sich offenbar

um eine Meta-Evaluation durch das Evaluationsteam selbst und somit um eine Bewertung ihrer eigenen Arbeit handelte.

Auch bei den Meta-Evaluationen von Caspari, bei welchen unter anderem die DeGEval-Standards eingesetzt wurden, fehlen Hinweise zur Eignung der Standards, deren Vollständigkeit oder aufgetretenen Widersprüchen (Caspari, 2010, 2011).

Ein anderes Forschungsprojekt beschäftigte sich mit der Anwendbarkeit der DeGEval-Standards im Kontext der Berufsbildung (Beywl & Speer, 2003, 2004). Mittels Literaturanalyse wurde hier gefunden, dass nicht jeder Standard in jedem einzelnen Projekt gleichermaßen verwendbar war, über alle Projekte hinweg jedoch alle Standards Anwendung fanden. Die Anforderungen der Genauigkeitsstandards erwiesen sich als gegensätzlich bzw. widersprüchlich zu denen der Nützlichkeitsstandards, da beide zugleich kaum eingehalten wurden und somit weitgehend unvereinbar waren. Die Autoren schlagen zwei ergänzende Evaluationsstandards vor, von denen einer die Qualitätsorientierung in der beruflichen Bildung betrifft, während der andere sich auf Evaluationen allgemein bezieht und die Forderung enthält, dass Evaluatoren das von ihnen verwendete Evaluationsmodell explizieren sollen. Zudem wird empfohlen, Begleitmaterial zu den Standards zu erarbeiten, welches diese für den Kontext der Berufsbildung konkretisiert.

Insgesamt kann der Forschungsstand zu den deutschen Evaluationsstandards als dürftig bezeichnet werden, da die hier dargestellten Ergebnisse auf sehr wenigen Studien basieren, in denen nur wenige Anwendungsgebiete der Evaluation vertreten sind. Die Forschung bleibt somit Antworten schuldig auf die Frage, ob die Standards umfassend sind, welche Standards einander widersprechen und wie mit solchen Widersprüchen in der Praxis umgegangen werden kann sowie ob die Erklärungen zu den Standards ausreichend und verständlich sind. Um die Qualität der Standards zu bestätigen bzw. zu deren Verbesserung beizutragen, sind Erfahrungen aus unterschiedlichen Praxisfeldern und aus den Perspektiven verschiedener Akteure notwendig. Ein diesbezüglicher Aufruf zur aktiven Diskussion der Standards sowie zu deren kritischer Bewertung mit dem Ziel ihrer Weiterentwicklung ist auch in den DeGEval-Standards selbst enthalten (DeGEval, 2008b).

2.2.5.5 Zusammenfassung des Forschungsstandes

Abschließend seien der Forschungsstand zu den Bereichen Meta-Evaluation und Evaluationsstandards sowie die diesbezüglichen Lücken und Forschungsbedarfe zusammengefasst, da sie den Ausgangspunkt für die im darauffolgenden Kapitel dargestellten Ziele und Fragestellungen der vorliegenden Arbeit bilden.

Im Bereich der Methodik von Meta-Evaluationen gab es in den letzten Jahren einige Bemühungen im anglo-amerikanischen Raum, allgemeine Ansätze zu erarbeiten sowie die Anwendung von Evaluationsstandards zu systematisieren. Zur Verwendung im Rahmen von Meta-Evaluationen werden die gängigen Standards für Evaluationen empfohlen, die ggf. durch weitere Richtlinien ergänzt werden können. Auf die Frage, wie die Standards zur Bewertung von Evaluationen eingesetzt werden können, bietet die Literatur bislang wenig Hilfestellung. In den Standards selbst sind kaum konkrete Hinweise enthalten und Studien, in denen sie bei Meta-Evaluationen eingesetzt wurden, erläutern ihr Vorgehen selten ausführlich genug, um eine Orientierung zu ermöglichen. Auch die Frage, ob der Einsatz von zwei Beurteilern der Einschätzung durch eine Person überlegen ist, lässt sich bisher nicht beantworten. Externe Evaluationsbewertungen verfügen oftmals nicht über alle Informationen, um die Einhaltung aller Standards verlässlich bewerten zu können.

Die Befunde zu den amerikanischen JCSEE-Standards lassen erkennen, dass Evaluationen in der Regel die *Accuracy*-Standards (Genauigkeit) und die *Propriety*-Richtlinien (Fairness) gut erfüllen, während im Bereich der *Utility* (Nützlichkeit) und *Feasibility* (Durchführbarkeit) mehr Verstöße vorkommen. Die Eignung der JCSEE-Standards für Meta-Evaluationen wurde lediglich in zwei Studien behandelt, deren Bewertung positiv ausfiel. Die teilweise geäußerte Kritik bezieht sich auf die Realitätsferne der JCSEE-Standards und damit auf einen Punkt, der auch an den deutschen DeGEval-Standards beanstandet wird.

Zu den DeGEval-Standards existiert sehr wenig Forschung. Aus den wenigen Publikationen wird deutlich, dass die deutschen Standards akzeptiert und überwiegend positiv bewertet werden. Diskutiert wird, ob die Allgemeinheit der Standardformulierungen angemessen ist oder ob spezifische Standards für die verschiedenen Evaluationsfelder angestrebt werden sollten, da diese eine höhere praktische Relevanz der Richtlinien bewirken könnten. Auch bezüglich der Ausführlichkeit der

DeGEval-Standards gibt es gegensätzliche Meinungen. So wird teils die derzeitige, eher knappe Version für ausreichend gehalten, teils aber auch eine deutlich umfangreichere Fassung analog zum amerikanischen Handbuch der JCSEE-Standards gefordert.

Die hier beschriebenen Befunde heben zugleich das Forschungsdefizit hervor, das in dem relativ jungen Forschungsgebiet der Evaluationsstandards und Meta-Evaluation herrscht. Bereits der angesprochene, geringe Umfang an Literatur zu diesem Gebiet verlangt weitere Forschungsaktivitäten zu diesen Themen. Die Zahl der deutschen Publikationen bleibt deutlich hinter jenen aus dem anglo-amerikanischen Raum zurück, wo die Thematik der Meta-Evaluation und der Evaluationsstandards bereits größere Relevanz besitzt. Insbesondere mit den DeGEval-Standards haben sich bislang nur einzelne Studien beschäftigt, sodass viele Forschungsfragen offen sind.

So wären detailliertere Berichte über die bei Meta-Evaluationen angewendete Methodik wünschenswert, da diese bislang häufig nur unzureichend geschildert wird. Dies hat zur Folge, dass das Entwicklungspotential solcher Studien ungenutzt bleibt. Neben einer ausführlicheren Beschreibung der Methodik wären auch Hinweise auf die damit gemachten Erfahrungen sowie daraus abgeleitete Empfehlungen hilfreich, um die Methodik von Meta-Evaluationen weiterentwickeln zu können. Des Weiteren bedarf es insbesondere im Hinblick auf die DeGEval-Standards solcher Untersuchungen, die deren Eignung für die Durchführung von Meta-Evaluationen erforschen. So wurde bislang nicht darauf eingegangen, ob diese Richtlinien umfassend genug sind, um alle für Meta-Evaluationen relevanten Bereiche abzudecken. Auch der Kritikpunkt der Realitätsferne könnte hier behandelt werden, um ggf. Entwicklungsbedarf und -möglichkeiten für die Standards abzuleiten. Zu diesem Zweck wären auch Berichte zu Erfahrungen mit den Standards sowie zu Herausforderungen der Evaluationspraxis hilfreich, da das Lernpotential bisheriger Erfahrungen bislang kaum genutzt wird. Es gilt auch zu klären, welche Standards einander in der Forschungsrealität widersprechen und wie mit dieser Problematik umgegangen wird, um anderen (Meta-)Evaluatoren Orientierungshilfen zu bieten.

Solche Forschungsaktivitäten würden der Forderung nach mehr Meta-Evaluationen nachkommen sowie die bisher fehlende inhaltliche Diskussion der Standards anregen. Sie würden zudem dem Aufruf der

Autoren der Evaluationsstandards Folge leisten, wonach die Richtlinien kritisch betrachtet und diskutiert werden sollen, um Ansätze für deren Weiterentwicklung zu finden.

3 Ziele und Fragestellungen

Anhand der im Theorieteil dargestellten Erkenntnisse wird deutlich, dass Evaluationen ein aktuelles und relevantes Thema darstellen, dessen weitere Erforschung sinnvoll und wichtig ist. Insbesondere zum Bereich der Meta-Evaluation und der Evaluationsstandards offenbart sich bisher ein Mangel an Forschungsarbeiten. Dieser ist als kritisch einzuschätzen, da Meta-Evaluation, also die Überprüfung und Bewertung von Evaluationsstudien, ein wichtiges Instrument zur Qualitätssicherung von Evaluationen ist und damit einen bedeutenden Beitrag zur Sicherstellung der Güte sowie zur Verbesserung der Evaluationsforschung leisten kann. Hier bedarf es weiterer Forschung, um die Methodik weiterzuentwickeln und die Rolle zu definieren, die Evaluationsstandards im Rahmen von Meta-Evaluationen spielen können.

Die vorliegende Arbeit hat zum Ziel, zur Schließung der beschriebenen Lücken im Forschungsstand zu Meta-Evaluation und Evaluationsstandards beizutragen. Zu diesem Zweck wird eine Evaluationsbewertung durchgeführt und damit der Forderung nach mehr Meta-Evaluationen nachgekommen.

Die zentrale Zielsetzung besteht darin, Herausforderungen nachzuzeichnen, auf die Evaluierende in der Praxis stoßen können und daraus Empfehlungen zu generieren. Dazu wird eine Evaluation systematisch anhand von Evaluationsstandards bewertet. Da es sich um eine sehr umfassende Evaluation mit vielen Beteiligten und umfangreichen Datenerhebungen handelt, sind zahlreiche Herausforderungen und entsprechend auch vielfältige Empfehlungen zu erwarten. Wenngleich eine Evaluationsbewertung durchgeführt wird, steht nicht das Bewertungsergebnis im Fokus, aus der (Nicht-)Einhaltung der einzelnen Standards sollen vielmehr Erkenntnisse zu folgenden Fragestellungen gewonnen werden:

- Welche Herausforderungen ergaben sich im Evaluationsprozess und wie wurde mit diesen umgegangen?
- Welche Empfehlungen können daraus abgeleitet werden?

Die gemachten Erfahrungen sollen somit für die Optimierung künftiger Evaluationen nutzbar gemacht werden, ein Lernprozess, der bislang viel zu selten realisiert wird. Da es sich bei der vorliegenden Arbeit um die Bewertung einer einzigen Evaluation handelt, besteht nicht der Anspruch, dass die abgeleiteten Empfehlungen für jede künftige Evaluationsstudie passend sein werden. Um dennoch ihren Geltungsbereich zu erhöhen, werden Befunde aus weiteren Studien miteinbezogen. Zudem ist zu hoffen, dass die hier ausgesprochenen Empfehlungen langfristig durch Erkenntnisse aus weiteren Meta-Evaluationen ergänzt werden.

Die Bewertung der Evaluation erfolgt anhand der deutschen Standards für Evaluation der DeGEval (2008b) und soll weiterhin Antworten auf folgende Fragestellungen erbringen:

- Stellen die Evaluationsstandards der DeGEval ein geeignetes Instrument für die Bewertung von Evaluationen dar?
- Sind diese Richtlinien umfassend, bilden also alle relevanten Aspekte ab?
- Findet der Kritikpunkt der Realitätsferne in Bezug auf die DeGEval-Standards Bestätigung oder Widerspruch?

In der zu bewertenden Evaluation wurden die Evaluationsstandards nicht explizit befolgt. Daher sind auch Aussagen dazu möglich, wie intuitiv diese Richtlinien sind.

- Welche Standards wurden intuitiv eingehalten und welche fanden weniger Beachtung?
- In Einklang mit dem Forschungsstand wird erwartet, dass die Genauigkeits- und Fairnessstandards am besten erfüllt wurden.

Ein weiterer Nutzen dieser Arbeit besteht darin, dass, im Gegensatz zu vielen anderen Publikationen, die Methode der Meta-Evaluation möglichst detailliert erläutert wird, um die Diskussion zur Methodik bei Meta-Evaluationen zu fördern und jene langfristig zu verbessern. Zu diesem Zweck wird die eingesetzte Methode zudem reflektiert und kritisch bewertet.

Die Meta-Evaluation wird als interne Meta-Evaluation realisiert, da die Autorin und Meta-Evaluatorin auch Mitglied des Evaluationsteams war. Damit stehen ihr für die Bewertung umfangreiche Daten, Unterlagen und Erfahrungen aus dem Evaluationsprozess zur Verfügung, was einen Vorteil der internen Evaluationsbewertung darstellt. Demgegenüber reichen bei externen Meta-Evaluationen die vorhandenen Informationen oftmals nicht für eine Bewertung der Einhaltung aller Evaluationsstandards aus. Interne Meta-Evaluationen können auch eine Grundlage für externe Meta-Evaluationen darstellen. Die vorliegende Arbeit kann somit als Ausgangspunkt für eine solche dienen, sowie den dadurch möglichen Vergleich der Bewertungen interner und externer Meta-Evaluatoren gestatten, welcher wiederum einen wertvollen Beitrag zur Methodikdiskussion leisten kann.

Des Weiteren hat diese Arbeit zum Ziel, die bislang mangelnde inhaltliche Diskussion zu beleben. Dies wird einerseits durch die genannten Fragestellungen verfolgt, andererseits soll ein Beitrag zu der umstrittenen Frage geleistet werden, ob die DeGEval-Standards in ihrer allgemeinen Formulierung angemessen sind oder ob feldspezifische Standards eine sinnvolle Zielsetzung darstellen. Zusätzlich wird zu dem weiteren Diskussionspunkt Stellung genommen, ob die deutschen Evaluationsstandards in ihrer bisherigen Form umfassend genug sind und ausreichende Informationen zur Verfügung stellen, oder ob eine ausführlichere Ausgabe in Form eines Handbuches empfehlenswert ist.

Insgesamt wird mit vorliegender Arbeit dem Aufruf der Autoren der Evaluationsstandards Folge geleistet, wonach die Standards kritisch betrachtet werden sollen. Sie verfolgt das Ziel, einerseits Erkenntnisse zu spezifischen Fragestellungen zu erbringen sowie andererseits die Diskussion in einem noch jungen Forschungsgebiet zu fördern.

4 Meta-Evaluation – Bewertung einer Evaluationsstudie

In diesem Kapitel wird die Meta-Evaluation vorgestellt, die auf der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ basiert und Antworten auf die im vorigen Abschnitt aufgeworfenen Fragestellungen liefern soll. Zunächst wird das Bundesprogramm „Elternchance ist Kinderchance“ beschrieben und anschließend dessen Evaluation dargestellt. In beiden Abschnitten werden die zum Verständnis der vorliegenden Arbeit notwendigen Informationen bereitgestellt, auf eine detaillierte oder gar vollständige Schilderung des Programms und der Evaluation wird hingegen verzichtet, da dies den Rahmen dieser Arbeit übersteigen würde. Der an weiteren Einzelheiten interessierte Leser sei stattdessen auf den Abschlussbericht der Evaluation (Müller et al., 2015) sowie auf den Nachbericht des Moduls 3 (Beck, Schwaß & Stemmler, 2016) verwiesen, denen die in Abschnitt 4.1 vorgestellten Angaben entnommen sind.

Im zweiten Abschnitt wird die Methode der Meta-Evaluation behandelt und nachfolgend die eigentliche Evaluationsbewertung dargelegt. Deren Ergebnisse werden abschließend zusammengefasst und im darauffolgenden Kapitel diskutiert.

4.1 Die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“

4.1.1 Das Bundesprogramm „Elternchance ist Kinderchance“

Das Bundesprogramm „Elternchance ist Kinderchance – Elternbegleitung der Bildungsverläufe der Kinder“ war ein vom Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) ins Leben gerufenes, präventiv ausgerichtetes Programm, das zum Ziel hatte, der sozialen Bildungsungleichheit in Deutschland entgegenzuwirken, indem Eltern als Bildungsvermittler verstanden und in dieser Rolle gestärkt wurden. Das Bundesprogramm, dessen Laufzeit im Jahr 2011 begann und bis Dezember 2014 dauerte, bestand aus zwei Bereichen. Dies waren die Qualifizierung von Fachkräften in der Familienbildung zu sogenannten *Elternbegleitern* einerseits und die Erprobung lokaler Vernetzungsstrukturen im Bereich der Eltern- und Familienbildung

(Programmbereich *Elternbegleitung Plus*) andererseits. Beide Programmbereiche verfolgten das Ziel, Eltern in ihren Erziehungs- und Bildungskompetenzen zu fördern und ihr Selbstverständnis als Bildungsvermittler für ihre Kinder zu stärken. Die Zielgruppe stellten Familien mit Kindern von der Geburt bis zum Ende der Grundschulzeit dar. Das Programm richtete sich dabei gezielt an bildungsferne Eltern, an Familien mit geringen sozio-ökonomischen Ressourcen und an Familien mit Migrationshintergrund, da es für deren Kinder in der Bildungslaufbahn besonders hohe Hürden zu überwinden gilt.

4.1.1.1 Qualifizierung von Fachkräften zu Elternbegleitern

Um die Ziele des Bundesprogramms zu erreichen, sollten die Eltern Unterstützung durch speziell geschulte Personen erhalten, welche ihnen z. B. als Ansprechpartner in Bildungs- und anderen Fragen zur Seite stehen und als Vertrauenspersonen fungieren sollten. Dies waren Fachkräfte der Familienbildung, die eine Weiterqualifizierung zum Elternbegleiter durchlaufen hatten. Die Qualifizierung sollte ihnen das entsprechende Wissen und die erforderlichen Kompetenzen vermitteln, um in der Praxis mit niedrigschwelligen Angeboten Familien erreichen, sie kompetent beraten und ihnen in Bildungsangelegenheiten weiterhelfen zu können. Auch die Vernetzung und Kooperation mit anderen Bildungsakteuren wurden thematisiert und interkulturelle Kompetenz ebenso wie diagnostische Fähigkeiten gestärkt.

Die Qualifizierungskurse wurden von zwei Trägern durchgeführt, der Bundesarbeitsgemeinschaft Familienbildung & Beratung e. V. (AGEF) und einem Trägerkonsortium, welches aus der Arbeitsgemeinschaft für katholische Familienbildung e. V. (AKF), dem Arbeiterwohlfahrt Bundesverband e. V. (AWO), der evangelischen arbeitsgemeinschaft familie e. V. (eaf), der Deutschen Evangelischen Arbeitsgemeinschaft für Erwachsenenbildung e. V. (DEAE), der Katholischen Bundesarbeitsgemeinschaft für Einrichtungen der Familienbildung (Kath. BAG) und dem Paritätischen Bildungswerk Bundesverband e. V. gebildet wurde. Beide Träger setzten die Qualifizierungen unabhängig voneinander um und hatten dazu jeweils ein eigenes, fachlich geprüftes Curriculum erstellt. Die Inhalte der beiden Curricula stimmten in wesentlichen Punkten überein, unterschieden sich aber in ihrer Schwerpunktsetzung sowie in der verwendeten Terminologie. Die Qualifizierungen wurden in Form von Präsenzveranstaltungen mit

120 (AGEF) bzw. 84 (Trägerkonsortium) Unterrichtseinheiten realisiert, kombiniert mit einem Selbststudium im Umfang von bis zu 120 Arbeitsstunden. Es wurden dabei keine konkreten Handlungsmodelle vermittelt, sondern vielmehr die Selbstreflexion der Fachkräfte und ihr professionelles Handlungsrepertoire gefördert. Dieses Ansetzen an den Ressourcen der einzelnen Teilnehmer, das Vorliegen zweier unterschiedlicher Curricula sowie die Tatsache, dass die Curricula den Dozenten Freiraum zur individuellen Ausgestaltung der Kurse ließen, führten dazu, dass die Qualifizierungen zu Elternbegleitern eine gewisse Heterogenität aufwiesen.

Die Teilnahme an der Qualifizierung war für die Fachkräfte kostenlos, da die Finanzierung vom BMFSFJ übernommen wurde. Im Verlauf der ersten Programmphase wurden bis Ende des Jahres 2014 bundesweit mehr als 5 500 Fachkräfte zu Elternbegleitern qualifiziert. Auch ab dem Jahr 2015 werden im Rahmen des Programms „Elternchance II – Familien früh für Bildung gewinnen“, welches eine Weiterentwicklung von „Elternchance ist Kinderchance“ darstellt, weiterhin Qualifizierungen durchgeführt.

4.1.1.2 Erprobung lokaler Vernetzungsstrukturen (Elternbegleitung Plus)

Im Rahmen des Programmbereichs *Elternbegleitung Plus* erhielten ab März 2012 bundesweit 100 Einrichtungen eine Förderung von bis zu 10 000 Euro jährlich. Diese Modellstandorte befanden sich in räumlicher Nähe zu einer „Schwerpunkt-Kita Sprache & Integration“ und in diesen Einrichtungen wirkten ein oder mehrere Elternbegleiter. Die finanzielle Unterstützung sollte den Modellstandorten den Aufbau lokal vernetzter Strukturen zur Bildungsbegleitung von Eltern ermöglichen und die Entwicklung und Erprobung von Angeboten für Familien sowie von neuen Zugängen zu diesen fördern. Im Rahmen der Vernetzung sollte den *Elternbegleitung Plus*-Einrichtungen die Rolle eines Vermittlers zukommen, der Eltern Unterstützungsmöglichkeiten aufzeigen und ihnen Angebote empfehlen kann. Da es sich bei der Förderung der Modellstandorte um eine Anschubfinanzierung handelte, endete dieser Programmbereich im Dezember 2014.

4.1.2 Rahmenbedingungen der Evaluation

Das Bundesprogramm „Elternchance ist Kinderchance“ wurde durch das Deutsche Jugendinstitut e. V. (DJI) in München und die Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) evaluiert. Dass das BMFSFJ die für eine umfangreiche, externe Evaluierung seines Programms erforderlichen Mittel bereitstellte, ist positiv hervorzuheben, insbesondere in Anbetracht der Tatsache, dass im Bereich sozialer Programme noch relativ selten Evaluationen realisiert werden (vgl. Abschnitt 2.1.5).

Die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ lässt sich anhand der im Theorieteil vorgestellten Charakteristika folgendermaßen einordnen (vgl. Abschnitt 2.1.3): Es handelte sich um eine externe Evaluation, die sowohl eine formative als auch eine summative Zielsetzung verfolgte, wie im nächsten Abschnitt erläutert wird. Den Evaluationsgegenstand stellte ein soziales Programm dar. Dieses wurde in Bezug auf die Güte seiner Implementierung sowie auf seine Wirksamkeit untersucht, wobei auch Optimierungsmöglichkeiten aufgezeigt wurden. Damit hatte die Evaluation eine Kontroll- und Optimierungsfunktion. Des Weiteren diente sie Legitimierungszwecken (vgl. Abschnitt 4.2.2.1.2) und brachte zusätzlich einen bedeutenden Erkenntnisgewinn mit sich. Bezüglich des Zeitpunkts der Evaluation kann diese einerseits eingeteilt werden als Evaluation der Programmdurchführung, da die Implementierung der Maßnahme unter Alltagsbedingungen erforscht wurde. Da zudem Veränderungen untersucht wurden, die sich aufgrund des Programms ergaben, handelte es sich weiterhin auch um eine Evaluation der Programmwirksamkeit. Schließlich wies die vorliegende Evaluation Merkmale sowohl des zielorientierten als auch des managementorientierten Ansatzes auf. Denn zum einen hatte sie die Überprüfung der Zielerreichung zum Ziel und sollte Optimierungsmöglichkeiten aufzeigen. Zum anderen orientierte sich die Evaluation an den Interessen des Auftraggebers, die es in erster Linie zu bedienen galt, und fand in enger Abstimmung mit diesem statt. Da letzterer Aspekt im Rahmen der Evaluationsbewertung (Abschnitt 4.2.2) mehrmals aufgegriffen wird, wird die vorliegende Evaluation im Folgenden als *auftraggeberorientierte Evaluation* (nach Bortz & Döring, 2006) bezeichnet.

Das Evaluationsprojekt begann im April 2012 und seine Beendigung war im Dezember 2014 angesetzt, was einer geplanten Laufzeit von fast

drei Jahren entsprach. Tatsächlich verzögerte sich jedoch die Abgabe des Abschlussberichts und damit das Ende der Evaluation, was nachfolgend näher erläutert wird. Die Evaluation war offen ausgeschrieben und im Vergabeverfahren erhielt die Bietergemeinschaft aus DJI und FAU Anfang April den Zuschlag. Die Rechte und Pflichten von Auftragnehmer und Auftraggeber wurden in Werkverträgen zwischen DJI und BMFSFJ und zwischen FAU und BMFSFJ einerseits, sowie in einer Kooperationsvereinbarung und einer Vereinbarung zum Datenaustausch zwischen DJI und FAU andererseits festgehalten. Das DJI war für die Evaluierung der Qualifizierung zum Elternbegleiter (Modul 1) und der Modellstandorte Elternbegleitung Plus (Modul 2) zuständig, die FAU erforschte die Wirkungen der Elternbegleitung auf Ebene der Eltern und Kinder (Modul 3). Für die Durchführung der Evaluation standen dem DJI von Mai 2012 bis Dezember 2014 vier Teilzeitstellen für wissenschaftliche Mitarbeiter (50 %) zur Verfügung, an der FAU waren im gleichen Zeitraum zwei Teilzeitstellen für wissenschaftliche Mitarbeiter (50 %) vorgesehen. Unterstützt wurden die Projektmitarbeiter durch vier (DJI) bzw. 15 (FAU) studentische und wissenschaftliche Hilfskräfte. Die Evaluation wurde von DJI und FAU in enger, vertrauensvoller Kooperation durchgeführt. Zu diesem Zweck wurden des Öfteren Video- oder Telefonkonferenzen zum Austausch im gesamten Team abgehalten, und auch Arbeitstreffen in München und Erlangen organisiert. Etwas beeinträchtigt wurden die Evaluationsaktivitäten durch die Personalfluktuation im Projekt, welche neben einem höheren Einarbeitungsaufwand auch Erfahrungsverlust mit sich brachte. So schieden beim DJI zwei der vier Mitarbeiterinnen mitten im Projekt aus und eine weitere gegen Ende der Evaluation, weshalb insgesamt zwei neue Mitarbeiterinnen hinzukamen. Auch bei der FAU verließ ein Mitarbeiter das Team und wurde durch eine neue Mitarbeiterin ersetzt. Zudem schied dort die Projektkoordinatorin (die an der Universität angestellt gewesen war) ersatzlos aus, sodass die koordinativen Aufgaben nun den beiden Projektmitarbeiterinnen zufielen. Neben jeweils einer Projektmitarbeiterin an DJI und FAU verblieben somit nur die beiden Projektleitungen, Prof. Dr. Sabine Walper (DJI) und Prof. Dr. Mark Stemmler (FAU) von Beginn bis Ende im Projekt.

Wie bereits erwähnt, stellte die vorliegende Evaluation eine auftraggeberorientierte Evaluation dar, was bedeutete, dass der Auftraggeber den zentralen Stakeholder darstellte, das Evaluationsteam vertraglich

an dessen Weisungen gebunden war, und die Evaluation darauf ausgelegt war, die Bedarfe des Auftraggebers zu erfüllen. Dies brachte eine enge Abstimmung der einzelnen Evaluationsschritte zwischen den Evaluatoren und dem BMFSFJ mit sich, wobei dieser Austausch (neben elektronischer und fernmündlicher Korrespondenz) anhand mehrerer Projekttreffen in Berlin erfolgte. Eine ähnlich enge Absprache gab es mit dem Kompetenzteam Wissenschaft, welches aus Wissenschaftlern bestand, eine beratende Funktion innehatte und selbst eng mit dem Bundesministerium zusammenarbeitete.

An der Evaluation waren zudem weitere Stakeholder beteiligt. Dazu gehörten die beiden bereits erwähnten Träger, die AGEF und das Trägerkonsortium, welche die Qualifizierung zum Elternbegleiter umsetzten. Auch die Dozenten beider Träger, die die Qualifizierungskurse durchführten, sowie die von ihnen ausgebildeten Elternbegleiter und auch die von diesen Fachkräften begleiteten Eltern stellten Stakeholdergruppen dar. Mit diesen Gruppen bestand insgesamt weniger Austausch als mit dem Auftraggeber, was in der Bewertung des Standards N1 nachgezeichnet wird (vgl. Abschnitt 4.2.2.1.1). Das Evaluationsteam bemühte sich jedoch insbesondere bei den Trägern, die in einem Konkurrenzverhältnis zueinander standen, um eine vertrauensvolle Zusammenarbeit, bei der kein Träger bevorzugt wurde. Daher wurden mit beiden Trägern Workshops veranstaltet, die dem Informationsaustausch und der Diskussion dienten. Zusätzlich besuchte mindestens ein Mitglied des Evaluationsteams die jährlichen Fachtagungen der Träger. Weitere Möglichkeiten zum Austausch ergaben sich im Rahmen von Workshops, die das BMFSFJ in verschiedenen Projektphasen in Berlin ausrichtete und an denen neben Auftraggeber, Kompetenzteam, Evaluatoren und Trägern auch zwei weitere Stakeholder teilnahmen: Die Impulsgruppe und die Servicestelle Elternchance. Die Impulsgruppe entsprach einem wissenschaftlichen Beirat, der sich aus Experten im Bereich der Evaluationsforschung sowie der Erwachsenenbildung und Familienforschung zusammensetzte und die Evaluation begleitete. Die Servicestelle Elternchance war für die fachlich-inhaltliche sowie finanziell-administrative Begleitung und Umsetzung des Programms zuständig. Neben der Beratung und Begleitung des Auftraggebers sowie der Zuwendungsempfänger erhob die Servicestelle im Rahmen eines Monitorings auch Daten bei Qualifizierungsteilnehmern sowie bei den Einrichtungen *Elternbegleitung Plus*, welche die Evaluatoren ebenfalls nutzen konnten.

Zwei weitere Veranstaltungen, die Erwähnung finden sollten, waren der 1. und 2. Bundeskongress Elternbegleitung, welche das BMFSFJ im Mai 2013 und im November 2014 in Berlin ausrichtete. Diese dienten dem fachlichen Austausch und es waren Elternbegleiter sowie Fachkräfte aus *Elternbegleitung Plus*-Einrichtungen ebenso wie Vertreter des Evaluationsteams, des Kompetenzteams und weiterer Stakeholder eingeladen.

Im Rahmen der Evaluation wurden verschiedene Berichte erstellt, deren Hauptadressat entsprechend der Ausrichtung der Evaluation jeweils das BMFSFJ war. Im Werkvertrag vereinbart waren halbjährliche Sachstandsberichte, ein Zwischenbericht zum 01.04.2013 und ein Abschlussbericht zum 31.12.2014. Die Sachstandsberichte sollten das BMFSFJ über die Aktivitäten des vergangenen halben Jahres und die Projektfortschritte informieren sowie die geplanten weiteren Schritte aufzeigen. Sie wurden jeweils pünktlich abgeliefert. Auch der Zwischenbericht (Walper et al., 2013), der fristgerecht eingereicht wurde, war lediglich für den Auftraggeber vorgesehen und wurde nicht veröffentlicht. Die Abgabe des Abschlussberichts verzögerte sich hingegen, was mehrere Gründe hatte. So erwiesen sich im Projektverlauf zusätzliche Auswertungen als notwendig, die der Auftraggeber für eine detaillierte und passgenaue Überarbeitung der beiden Ausbildungscurricula für deren Anwendung in der zweiten Phase des Bundesprogramms „Elternchance II“ benötigte (vgl. Abschnitt 4.2.2.1.2). Des Weiteren äußerte der Auftraggeber relativ spät (im April 2014) den Wunsch nach einer kurzen Broschüre mit Erscheinungstermin November 2014 (rechtzeitig zum 2. Bundeskongress Elternbegleitung), in welcher die Evaluatoren die zentralen Ergebnisse in verständlicher und anschaulicher Form aufzeigten und die an ein breiteres Publikum verteilt werden konnte als der wissenschaftliche Abschussbericht. Ziel dieser Broschüre war es, Elternbegleiter und eine breite Fachöffentlichkeit zu erreichen sowie innerhalb des Ministeriums für die Fortsetzung des Bundesprogramms zu werben. Diese Kurzbroschüre (Gerleigner et al., 2014) wurde im November 2014 als Internetpublikation veröffentlicht und konnte auch als Printversion bestellt werden. Zwar vereinbarten BMFSFJ und Evaluationsteam aufgrund dieser zusätzlichen Aufgaben, andere Leistungen wegzulassen (z. B. einzelne Erhebungen und einige Auswertungen), dennoch war abzusehen, dass der ursprüngliche Abgabetermin des Abschlussberichts nicht eingehalten

werden konnte. Denn auch die Datenerhebungen hatten sich als sehr aufwendig erwiesen und deutlich mehr Ressourcen erfordert als geplant. Insbesondere bei der Kontrollgruppe war ein schlechter Rücklauf zu verzeichnen, was mit erhöhtem Arbeitsaufwand verbunden war (vgl. Abschnitt 4.1.3.2.1). Daher wurde der Abgabetermin des Abschlussberichts einvernehmlich auf Ende Januar 2015 verschoben und nach einigen Korrekturdurchgängen wurde der Bericht Anfang August 2015 veröffentlicht (Müller et al., 2015). Zudem wurden sowohl für die Erhebung als auch für die detaillierte Auswertung der Kontrollgruppendaten im Modul 3 zusätzliche Ressourcen benötigt. Das BMFSFJ stellte dafür weitere Mittel bereit und kam mit den Evaluatoren der FAU überein, die Kontrollgruppendaten aus dem Abschlussbericht auszuklammern und stattdessen einen eigenen Nachbericht mit diesen Daten zu verfassen. Dieser wurde im Juni 2015 abgegeben und Anfang des Jahres 2016 veröffentlicht (Beck et al., 2016). Alle zur Veröffentlichung vorgesehenen Berichte (Kurzbroschüre, Abschlussbericht, Kontrollgruppen-Nachbericht) wurden vor der Publikation stets mit dem BMFSFJ und dem Kompetenzteam besprochen, nach deren Rückmeldung ggf. durch das Evaluationsteam überarbeitet und schließlich vom Auftraggeber freigegeben.

4.1.3 Ziele, Design und Methoden der Evaluation

Die hier vorgenommene Schilderung der Evaluationsziele, des Evaluationsdesigns und der verwendeten Methoden soll dem Leser einen Überblick über die methodische Anlage und Durchführung der Evaluation vermitteln. Für detailliertere Informationen sei wiederum auf den wissenschaftlichen Abschlussbericht der Evaluation verwiesen (Müller et al., 2015).

4.1.3.1 Evaluationsziele

Im Rahmen der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ sollte untersucht werden, wie die Implementierung der Maßnahme gelang und welche Optimierungsmöglichkeiten sich hierzu offenbarten (formative Evaluation). Zudem waren auch die Wirkungen des Programms von Interesse (summative Evaluation). Diese beiden Evaluationsziele waren mit folgenden Fragestellungen verbunden:

- **Inwieweit wurde die Qualifizierung zu Elternbegleitern erfolgreich umgesetzt?** Es sollte untersucht werden, ob die Fachkräfte wie vorgesehen für die Qualifizierung gewonnen werden konnten und ob die Teilnehmenden diese im Hinblick auf die Qualifizierungsziele sowie entsprechende Anforderungen in der Praxis als hilfreich und wirksam einschätzten.
- **Inwieweit ging die Qualifizierung mit den angestrebten Veränderungen in der Elternarbeit der Fachkräfte einher?** Hier galt es zu klären, welchen Lern- und Kompetenzzuwachs die Fachkräfte in der Qualifizierung erfuhren, inwieweit sich ihr professionelles Selbstverständnis als Elternbegleiter veränderte, ob sie neue Zugänge zu bildungsfernen Familien entwickelten und unter welchen Bedingungen sie die familiale Bildungsbegleitung in ihrem beruflichen Alltag umsetzen konnten.
- **Wie bewährte sich die Arbeit der im Rahmen des Programmbereichs *Elternbegleitung Plus* geförderten Modellstandorte?** In diesem Zusammenhang wurde untersucht, ob wie beabsichtigt neue, auf die Bildungsbegleitung von Familien ausgerichtete Netzwerkstrukturen im Sozialraum entstanden und welche Angebote und Zugänge besonders geeignet waren, um vermehrt bildungsferne Familien zu erreichen.
- **Inwieweit zeigte die Elternbegleitung die angestrebten Wirkungen auf Seiten der Eltern und Kinder?** Schließlich stellte sich die Frage, ob sich Eltern durch die Elternbegleitung in ihren Bildungskompetenzen bestärkt sahen, ob sie ihren Kindern vermehrt Lern- und Bildungsanregungen vermittelten und inwieweit die Kinder von den familialen Anregungsstrukturen profitierten.

4.1.3.2 Evaluationsdesign und Methodik

Zur Beantwortung der Evaluationsfragestellungen wurden umfangreiche Datenerhebungen durchgeführt, die quantitative sowie qualitative Befragungen umfassten. Die Evaluation war dabei in drei Module unterteilt: Modul 1 befasste sich mit den Elternbegleitern, in Modul 2 wurden die Modellstandorte untersucht und in Modul 3 wurde die Ebene der Eltern und Kinder erforscht (vgl. Abbildung 4 auf Seite 87). Für die Module 1 und 2 war das DJI zuständig, die Umsetzung des

Moduls 3 erfolgte durch die FAU (wobei es, der kooperativen Durchführung dieser Evaluation gemäß, zwischen DJI und FAU einen regen Austausch bezüglich der einzelnen Module gab).

Im Folgenden werden die drei Module der Evaluation einzeln näher beschrieben. Allen drei Modulen gemeinsam war die Durchführung einer Pilotstudie sowohl im quantitativen als auch im qualitativen Bereich, bei der die Fragebögen und Leitfäden auf Verständlichkeit und die Erhebung auf Durchführbarkeit hin getestet wurden (vgl. Müller et al., 2015; Walper et al., 2013). In allen Modulen wurde zudem darauf geachtet, die Befragungen ökonomisch zu gestalten, um den Aufwand für die Befragten so gering wie möglich zu halten. Eine weitere Anmerkung, die sich auf alle drei Module bezieht, betrifft das Auswertungsvorgehen bei den quantitativen Daten. Diese wurden stets mittels gängiger multivariater Analyseverfahren ausgewertet, deren Auswahl sich nach dem Skalenniveau der Daten richtete. Waren notwendige Voraussetzungen für bestimmte Verfahren nicht gegeben, wurde auf geeignete Verfahren ausgewichen. A priori wurde für alle inferenzstatistischen Signifikanzprüfungen ein Alpha-Niveau von .05 festgesetzt.

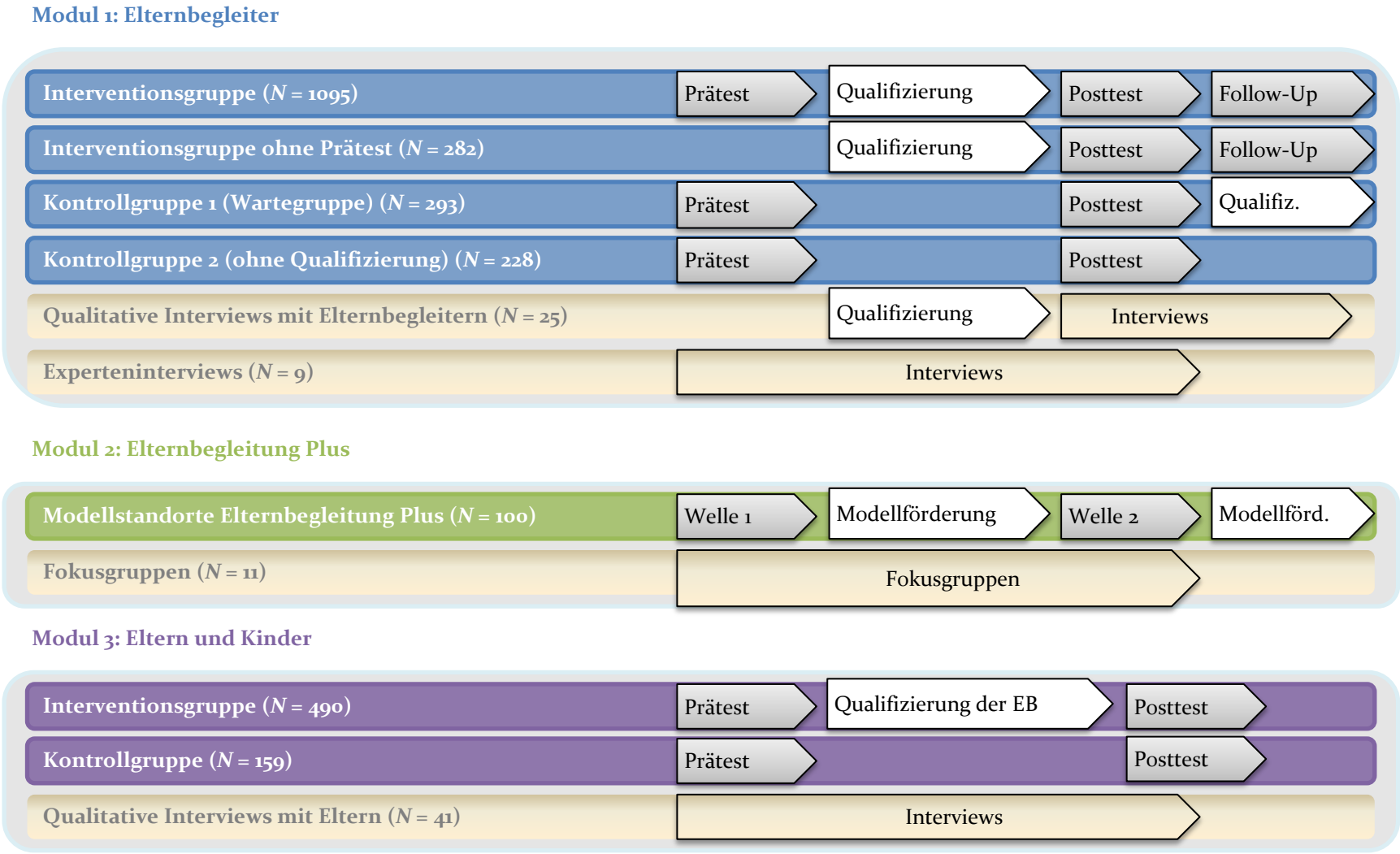


Abbildung 4: Evaluationsdesign
(nach Müller et al., 2015, S. 44)

4.1.3.2.1 Modul 1 – Elternbegleiter

Wie aus Abbildung 4 ersichtlich wird, wurden in jedem Modul quantitative und qualitative Daten erhoben. Im Modul 1 erfolgten die quantitativen Erhebungen anhand eines Prätest-Posttest-Follow-Up-Kontrollgruppenplans. Dabei fand der Prätest kurz vor Beginn der Qualifizierung statt, der Posttest nach deren Beendigung und die Follow-Up-Erhebung vier bis neun Monate nach Ende der Qualifizierung. Neben der Interventionsgruppe 1, die nach diesem Schema befragt wurde, gab es eine zweite Interventionsgruppe, die nur den Posttest und das Follow-Up durchlief. Somit konnte überprüft werden, ob die Teilnahme am Prätest sich auf das Antwortverhalten auswirkte. Es wurden zudem zwei Kontrollgruppen befragt, wobei die Kontrollgruppe 1 bzw. die Wartegruppe aus Fachkräften bestand, die planten, noch an der Qualifizierung teilzunehmen, und die Kontrollgruppe 2 sich aus Fachkräften zusammensetzte, bei denen dies nicht der Fall war. Beide Gruppen wurden mit dem Prätest und dem Posttest befragt, wobei der zeitliche Abstand dazwischen vergleichbar zu dem der Interventionsgruppe gewählt wurde. Die Wartegruppe sollte gegenüber der zweiten Kontrollgruppe den zusätzlichen Vorteil haben, dass bei diesen Befragten von vergleichbaren motivationalen, qualifikatorischen und institutionellen Voraussetzungen wie bei der Interventionsgruppe auszugehen war. Die Umfänge der vier Stichproben zu den verschiedenen Testzeitpunkten können Tabelle 2 entnommen werden, welche auch die Ausschöpfungsquoten enthält. Es ist zu beachten, dass in beiden Interventionsgruppen und in der Wartegruppe jeweils Fachkräfte, die durch die AGEF oder durch das Trägerkonsortium qualifiziert wurden, gleichermaßen enthalten waren. Auch Fachkräfte sowohl mit als auch ohne Anbindung an einen Modellstandort *Elternbegleitung Plus* wurden hier berücksichtigt.

Tabelle 2: Stichprobenumfänge und Ausschöpfungsquoten der drei Module

		Modul 1				**Modul 2**	**Modul 3**	
		Interventions-gruppe	Interventions-gruppe 2 ohne Prätest	Kontroll-gruppe 1 bzw. Wartegruppe	Kontroll-gruppe 2	Modellstand-orte *Elternbe-gleitung Plus*	Interven-tionsgruppe Eltern	Kontroll-gruppe Eltern
Erhebungszeitraum		11/2012 – 05/2014	02/2013 – 04/2014	11/2013 – 05/2014	05/2013 – 04/2014	12/2012 – 03/2014	01/2013 – 06/2014	05/2014 – 11/2014
Qualifizierte Fachkräfte bzw. geförderte Modellstandorte im äquivalenten Zeitraum		2096	599	1451[a]	–	100	–	–
Bruttostichprobe[b]		2055	650	390	1060	100	666	196
Bruttostichprobe (bereinigt)[c]		2036	644	388	1008	100	629	188
Nettostichprobe[d]	Prätest	1095	–	293	228	100	490	159
	Posttest	609	282	67	97	96	337	120
	Follow-Up	514	165	–	–	–	–	–
Ausschöpfungs-quote[e]	Prätest	53.8 %	–	75.5 %	22.6 %	100 %	77.9 %	84.6 %
	Posttest	29.9 %	43.8 %	17.3 %	9.6 %	96 %	53.6 %	63.8%
	Follow-Up	25.3 %	25.6 %	–	–	–	–	–
Rücklaufquote im Längsschnitt	T2 zu T1	55.6 %	–	22.9 %	42.5 %	96 %	68.8 %	75.5 %
	T3 zu T2	84.4 %	58.5 %	–	–	–	–	–
Gesamtausschöpfung[f]		25.3 %	25.6 %	17.3 %	9.6 %	96 %	53.6 %	63.8%

Anmerkungen. [a] Hochrechnung auf Basis der Teilnahmelisten der Qualifizierungsträger. [b] Bruttostichprobe = angeschriebene Fachkräfte und Modellstandorte bzw. Anzahl der von Fachkräften vermittelten Eltern. [c] Bruttostichprobe abzüglich stichprobenneutraler Ausfälle. [d] Nettostichprobe = Teilnehmende an der Befragung bzw. auswertbare Interviews. [e] Ausschöpfungsquote = Nettostichprobe/bereinigte Bruttostichprobe. [f] Gesamtausschöpfung = Ausschöpfungsquote, jeweils bezogen auf die letzte Erhebungswelle.

(nach Müller et al., 2015, S. 51)

Zur Rekrutierung der beiden Interventionsgruppen sowie der Wartegruppe wurden Fachkräfte, die sich für die Qualifizierung angemeldet hatten, per E-Mail um Teilnahme an der Befragung gebeten. Die Vermittlung der dazu notwendigen Kontaktinformationen erfolgte über die beiden Qualifizierungsträger, welche die Evaluation dadurch sehr unterstützten. Bei der Kontrollgruppe 2 gestaltete sich das Gewinnen von Befragungsteilnehmern als schwierig, was sich in sehr mäßigen Ausschöpfungsquoten widerspiegelte (vgl. Tabelle 2). Der Zugang zu den potentiellen Teilnehmern erfolgte hier über Einrichtungen, welche vom Evaluationsteam per E-Mail oder Brief angeschrieben wurden. Die Einrichtungen wurden deutschlandweit recherchiert, wobei hinsichtlich verschiedener Kriterien (Bundesland, Einrichtungsart und -größe sowie Träger) eine Übereinstimmung mit den Merkmalen der Interventionsgruppe angestrebt wurde. Die geringen Teilnehmerzahlen konnten auch durch eine zweite Rekrutierungswelle (Anschreiben weiterer recherchierter Einrichtungen), durch Erinnerungen der Fachkräfte (per E-Mail und teilweise auch telefonisch) sowie durch den Einsatz von Incentives nicht gesteigert werden. Da die Stichprobe der Kontrollgruppe 2 somit kleiner ausfiel als vorgesehen und ihre Erfassung zudem mehr Zeit und Ressourcen beanspruchte, konnte bei dieser Gruppe die ursprünglich geplante Follow-Up-Erhebung nicht realisiert werden.

Die quantitativen Erhebungen des Moduls 1 erfolgten mittels standardisierter Online-Fragebögen, die erhobenen Daten beruhten auf Selbstauskünften der Befragten. In den Fragebögen kamen neben etablierten Skalen auch eigens entwickelte Indikatoren zur Anwendung und sie enthielten ergänzend zu standardisierten Items auch offene Fragen. Die Befragungen sollten einerseits Informationen zu Inanspruchnahme und Akzeptanz der Qualifizierung liefern, aus denen Möglichkeiten der Nachsteuerung abgeleitet werden konnten (formative Evaluation). Andererseits sollten sie Daten für die Beantwortung der summativen Evaluationsfrage bereitstellen, inwieweit die Qualifizierung zu Wissens- und Kompetenzzuwachs sowie zu Veränderungen in der beruflichen Praxis der Elternbegleiter führte. Je nach zu beantwortender Fragestellung wurden einige Informationen zu mehreren Messzeitpunkten erhoben (beispielsweise Kompetenzen, um Kompetenzveränderungen abbilden zu können), andere hingegen nur einmalig erfasst (z. B. soziodemographische Merkmale). Die genauen Fragebogeninhalte (Items und Skalen zum jeweiligen Erhebungszeitpunkt)

können dem Abschlussbericht entnommen werden (Müller et al., 2015), an dieser Stelle seien zur Veranschaulichung lediglich die übergeordneten Inhaltsbereiche genannt:

- Teilnahme an der Qualifizierung, Stand der Qualifizierung
- Beschreibung und Bewertung der Qualifizierung
- Kenntnisse und Kompetenzen
- Familienbildung und -beratung, Zusammenarbeit mit Eltern
- Zielgruppen
- Professionelles Selbstverständnis
- Einstellungen
- Nachhaltigkeit
- Kooperation und Vernetzung
- Soziodemographie
- Bildung/Qualifikation
- Berufliche Situation
- Einwilligungen

Die quantitativen Erhebungen wurden durch qualitative Interviews mit Experten einerseits und mit Elternbegleitern andererseits ergänzt (vgl. Abbildung 4). Die leitfadengestützten Experteninterviews wurden mit insgesamt neun Vertretern der beiden Qualifizierungsträger geführt und hatten zum Ziel, relevantes Kontextwissen zur Umsetzung der Qualifizierungskurse zu erhalten. Dabei wurden unter anderem die Ziele der Befragten thematisiert, die konkrete Ausgestaltung der Lernprozesse in den Kursen besprochen und Erfahrungen erfragt. Die face-to-face-Interviews dauerten zwischen 60 und 90 Minuten, wurden transkribiert und relevante Inhalte wurden in Fallprotokollen zusammengefasst.

Die qualitativen Erhebungen bei den Elternbegleitern erfolgten in Form von teilnarrativen Interviews mit 25 Fachkräften (23 Frauen und zwei Männern). Dabei sollten unter anderem intendierte und nicht intendierte Effekte der Qualifizierung zur Sprache kommen sowie förderliche und hinderliche Bedingungen der Elternbegleitung erörtert werden. Die Teilnehmer an den teilnarrativen Interviews wurden mittels eines qualitativen Stichprobenplans ausgewählt, bei dem folgende Kriterien berücksichtigt wurden: Arbeitsfeld der Fachkräfte, Qualifizierungsträger und Zufriedenheit mit der Qualifizierung. Dadurch sollten verschiedene Arbeitsfelder sowie durch beide Träger

qualifizierte Personen ausgewogen repräsentiert und zudem sowohl sehr zufriedene als auch weniger zufriedene Elternbegleiter in der Stichprobe enthalten sein. Des Weiteren wurde auf Heterogenität hinsichtlich des Kursbeginns und -ortes, des Bundeslandes sowie des Alters der Befragten geachtet. Von den durchschnittlich eine Stunde dauernden Interviews wurden 11 face-to-face-Interviews realisiert, die übrigen 14 Gespräche aus Gründen der Kosten- und Zeitersparnis telefonisch geführt. Die Interviews wurden transkribiert und anhand qualitativer Kodierverfahren sowie mittels der dokumentarischen Methode ausgewertet.

4.1.3.2.2 Modul 2 – Modellstandorte Elternbegleitung Plus

Im Rahmen der quantitativen Erhebungen in Modul 2 wurden zu zwei Messzeitpunkten Daten in den geförderten Einrichtungen erhoben (vgl. Abbildung 4). Es handelte sich dabei nicht um eine klassische Vorher-Nachher-Testung, da die Modellförderung bereits vor dem Start der Evaluation begonnen hatte und auch erst nach Abschluss der Erhebungen endete. Da in diesem Modul keine summative Evaluation stattfand, wurde zudem keine Kontrollgruppe befragt. An der ersten Befragungswelle nahmen alle 100 geförderten Einrichtungen teil (Vollerhebung), an der zweiten Befragung, die ein Jahr später durchgeführt wurde, beteiligten sich 96 Modellstandorte (vgl. Tabelle 2). Befragt wurden jeweils die Einrichtungsleitungen.

Auch in Modul 2 kamen bei der quantitativen Erhebung Online-Fragebögen zum Einsatz, die durch die Einrichtungsleitungen ausgefüllt wurden, sodass es sich auch hier um Selbstberichtsdaten handelte. Anhand der Fragebögen wurde erfasst, wie in der Förderphase lokale Kooperations- und Vernetzungsstrukturen aufgebaut bzw. erweitert und genutzt, welche neuen zielgruppenspezifischen Angebote entwickelt und erprobt wurden und welche Zugangsweisen sich im Bereich der Bildungsbegleitung von Familien bewährten. Bei den Erhebungen in den *Elternbegleitung Plus*-Einrichtungen galt es zwei Aspekte zu beachten: Zum einen wurden an diesen Modellstandorten im Rahmen des Monitorings der Servicestelle Elternchance viele Daten zur Umsetzung der geplanten Modellprojekte erhoben, welche dem Evaluationsteam zur Verfügung gestellt werden konnten. Zum anderen hatten die Einrichtungen ein hohes Maß an Verwaltungsaufgaben zu bewältigen und wollten durch die Evaluation möglichst wenig belastet

werden. Daher wurden die Online-Fragebögen so knapp wie möglich gehalten. Zum ersten Befragungszeitpunkt wurde die Ausgangssituation hinsichtlich der institutionellen Rahmenbedingungen sowie relevanter Veränderungsbereiche erfasst. Dabei wurden unter anderem die Art und Zahl der Kooperationspartner, die Zufriedenheit mit der Kooperation, derzeitige und geplante Zugänge zu Familien sowie Formen und Themen der Angebote erfragt. Bei der zweiten Erhebungswelle wurden die Kooperation und Vernetzung im Rahmen der geförderten Angebote vertieft erfasst. Die Inhaltsbereiche der Online-Befragungen werden im Folgenden aufgelistet, detailliertere Informationen sind im Abschlussbericht der Evaluation (Müller et al., 2015) enthalten.

- Allgemeine Fragen zur geförderten Einrichtung
- Struktur der Einrichtung
- Beschreibung des Standortes
- Angebotsspektrum der Einrichtung
- Zielgruppen
- Kooperation und Vernetzung der Einrichtung
- Geförderte Angebote *Elternbegleitung Plus*
- Fragen zur ausfüllenden Person

Es wurden zudem qualitative Daten erhoben, indem an 11 Modellstandorten Fokusgruppen durchgeführt wurden (vgl. Abbildung 4), also themenzentrierte Gruppendiskussionen mit mehreren Akteuren eines Kooperationsnetzwerkes. Diese hatten zum Ziel, die Kooperation und Vernetzung im Sozialraum abzubilden und dabei sowohl Chancen als auch Grenzen zu erfassen. Dazu wurden zunächst Modellstandorte ausgewählt, die an den Fokusgruppen teilnehmen sollten. Die gewählten Einrichtungen sollten ein möglichst breites Spektrum der heterogenen Modellstandorte abdecken, wobei sich die Auswahl an folgenden Kriterien orientierte: Art und Größe der Einrichtung, Standort und Einzugsgebiet der Einrichtung, Anzahl der Netzwerke sowie Hinweise auf innovative Zugänge und/oder Angebote. Die ausgewählten Einrichtungen wurden gebeten, für die Fokusgruppe relevante Kooperationspartner zu benennen, die für sie große Bedeutung hinsichtlich der Entwicklung von Familienbildungsangeboten und dem Zugang zu Zielgruppen hatten. Somit fingen die Fokusgruppen die Perspektiven verschiedener Akteure vor Ort ein. Sie wurden von einer

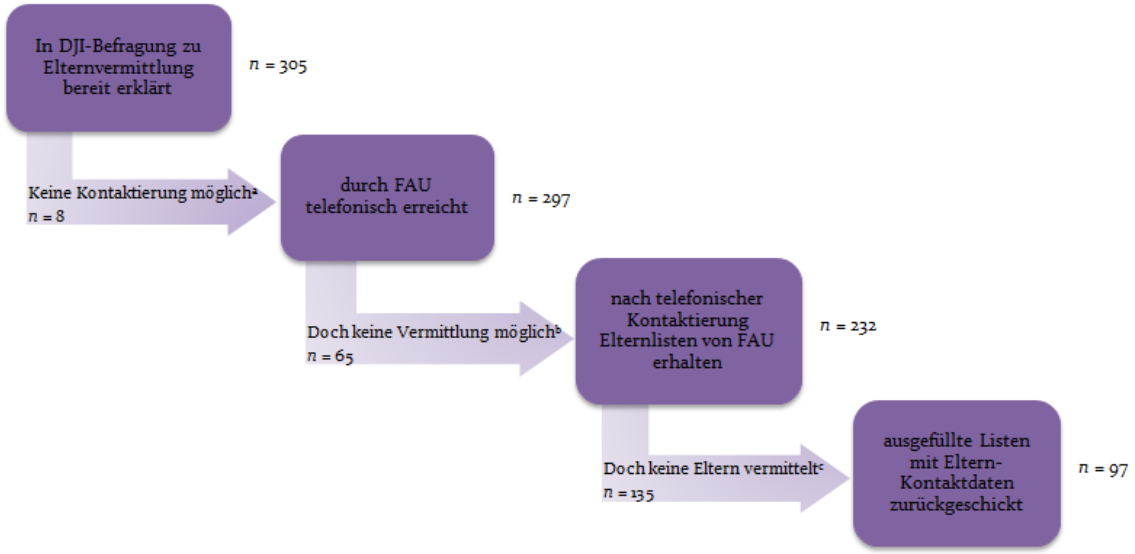
Moderatorin anhand eines offenen, diskussionsgenerierenden Leitfadens durchgeführt. Dieser umfasste unter anderem die folgenden Aspekte: Entwicklung der Kooperation, Angebote des Modellstandortes und Bewertung der Förderung. Die Fokusgruppen wurden transkribiert und mittels inhaltlich strukturierender qualitativer Inhaltsanalyse ausgewertet. Weiterhin wurden Kooperationsnetzwerke rekonstruiert und zudem die qualitativen Daten mit den Befunden der Onlinebefragungen im Sinne der Triangulation zusammengeführt.

4.1.3.2.3 Modul 3 – Eltern und Kinder

Modul 3 beschäftigte sich mit der Frage, inwieweit die Maßnahmen des Bundesprogramms bei den Familien als Zielgruppe ankamen. Befragt wurden dabei Eltern, Daten über die Kinder wurden nur indirekt über die Eltern erfasst. Die quantitative Datenerhebung erfolgte anhand eines Prätest-Posttest-Kontrollgruppendesigns (vgl. Abbildung 4), der Posttest fand dabei etwa ein halbes Jahr nach dem Prätest statt. Die Interventionsgruppe bestand aus Eltern, die durch Elternbegleiter unterstützt wurden und schloss auch Eltern mit ein, deren Elternbegleiter eine Anbindung an einen Modellstandort *Elternbegleitung Plus* hatten. Die Kontrollgruppe wurde aus Eltern gebildet, die keine Elternbegleitung erfuhren, aber eine Anbindung an Fachkräfte hatten, die nicht die Weiterqualifikation zum Elternbegleiter durchlaufen hatte. Die Stichprobenumfänge und Ausschöpfungsquoten beider Gruppen können Tabelle 2 entnommen werden.

Die an der Befragung teilnehmenden Eltern wurden über ihren Elternbegleiter bzw. über ihre Kontrollgruppen-Fachkraft vermittelt. Die Fachkräfte selbst konnten in der Online-Befragung von Modul 1 angeben, ob sie zur Vermittlung von Eltern, mit denen sie zusammenarbeiteten, bereit wären. Mit den Fachkräften, die dies bejahten, setzte sich das Evaluationsteam daraufhin telefonisch in Verbindung, erklärte Details zur Elternbefragung und in der Folge gab der Großteil (vgl. Abbildung 5) der Fachkräfte an, Eltern auf die Teilnahme an der Befragung ansprechen zu wollen. Diesen wurden postalisch Listen zugesendet, in welche sich die teilnahmebereiten Eltern mitsamt Kontaktdaten eintragen konnten. Dies hatte den Vorteil, dass die Fachkräfte nur als Vermittler fungierten und die Ansprache der Eltern übernahmen, die Eltern ihre Kontaktdaten jedoch eigenständig herausgaben. Dieses datenschutzkonforme Vorgehen wurde von den

Fachkräften begrüßt. Die Listen mit den Elterndaten wurden von den Fachkräften an die Evaluatoren zurückgeschickt, welche nachfolgend mit der Kontaktierung der Eltern begannen und die quantitative Befragung durchführten. Abbildung 5 veranschaulicht den Rücklauf dieser Rekrutierungsstrategie sowie Gründe für Drop-Outs für die Interventionsgruppe.

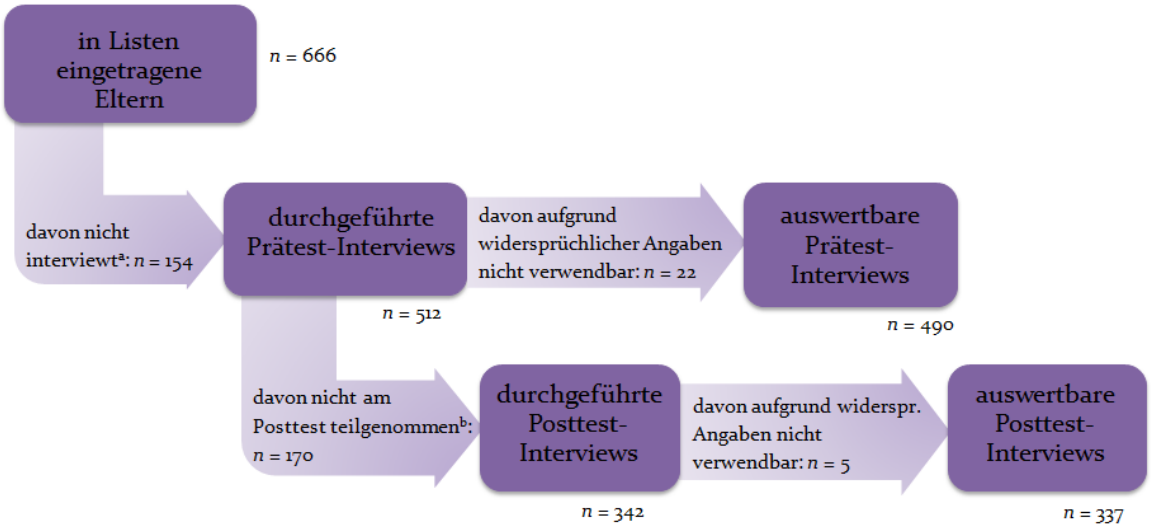


[a] Gründe: telefonisch nicht erreichbar; Telefonnummer falsch; keine Reaktion auf Kontaktierung per E-Mail
[b] Gründe: derzeit [noch] nicht als Elternbegleiter tätig; Qualifikation abgebrochen; Elternansprache abgelehnt, aus Zeitgründen
oder um die Vertrauensbeziehung zu den Eltern nicht zu gefährden; Kinder der Eltern nicht im passenden Alter
[c] Gründe: keine Eltern für Teilnahme gefunden; genauer Grund meist unklar

Abbildung 5: Drop-Out-Raten sowie -Gründe im Rahmen der Rekrutierung der Elternstichprobe durch die Fachkräfte (Modul 3, Interventionsgruppe)
(nach Müller et al., 2015, S. 53)

Aus Abbildung 5 wird ersichtlich, dass die zur Teilnahme vermittelten Eltern durch den Rekrutierungsprozess über die Fachkräfte bereits vorselektiert waren, da die Fachkräfte selbst entschieden, ob und welche Eltern sie vermittelten. Zudem wurden von weniger als der Hälfte der Fachkräfte, die Listen geschickt bekommen hatten, ausgefüllte Listen zurückgesendet. Von den letztendlich vermittelten Eltern konnten jedoch längst nicht alle interviewt werden, wie Abbildung 6 für die Interventionsgruppe zeigt. Auch hier sind Selektionsprozesse zu vermuten, da z. B. die schlechte Erreichbarkeit vieler Eltern mit einer schwierigen Situation dieser Familien in Verbindung stehen könnte,

welche die leicht erreichbaren Eltern nicht aufwiesen. Somit ist festzuhalten, dass die Elternstichprobe, welche die Daten für die Evaluation lieferte, eine selektive Stichprobe darstellte, die nur eingeschränkt repräsentativ für die Gesamtpopulation aller von Fachkräften der Familienbildung betreuten Eltern sein dürfte.



[a] Gründe: Proband (Pb) nicht erreichbar; Pb nach [teils mehreren] verpassten Interviewterminen nicht mehr erreicht; Pb gab an, keine Zeit zu haben; Deutschkenntnisse nicht ausreichend und keine Interviewer in gewünschter Sprache verfügbar; Kinder nicht im passenden Alter
[b] Gründe: aufgrund späten Prätests war Posttest-Erhebung zeitlich nicht mehr möglich; Pb nicht erreichbar; Pb nach [teils mehreren] verpassten Interviewterminen nicht mehr erreicht; Pb gab an, keine Zeit zu haben; Pb hatte im Prätest nicht in erneute Kontaktierung eingewilligt; Pb hatte weder zum Prä- noch zum Posttestzeitpunkt Kontakt zum Elternbegleiter

Abbildung 6: Drop-Out-Raten sowie -Gründe im Rahmen der Rekrutierung der Elternstichprobe (Modul 3, Interventionsgruppe)
(nach Müller et al., 2015, S. 54)

Wie bereits erwähnt, war die Rücklaufquote bei der Kontrollgruppe von Modul 1 sehr gering, was in erhöhtem Aufwand für die Rekrutierung sowie in einer deutlich längeren Dauer derselben resultierte (Abschnitt 4.1.3.2.1). Hier erwies es sich als besonders problematisch, dass die Elternbefragungen denen der Fachkräfte stets zeitlich nachgeschaltet waren (da die Fachkräfte erst nach Teilnahme an der Online-Befragung kontaktiert werden konnten und bis zur tatsächlichen Vermittlung von Eltern meist mindestens vier bis zwölf weitere Wochen vergingen). Der Posttest der Elternbefragung konnte dadurch erst so spät begonnen werden, dass die Kontrollgruppendaten bis zur Erstellung des Abschlussberichts der Evaluation nicht vollständig

vorlagen und in diesem nicht dargestellt werden konnten. In einem gesonderten Bericht wurden die vergleichenden Analysen, die unter Einbezug der Kontrollgruppe möglich waren (summative Evaluation), nachgeliefert (vgl. Abschnitt 4.1.2).

In Modul 3 fanden die quantitativen Befragungen mittels computergestützter Telefoninterviews statt. Diese wurden von geschulten Interviewern durchgeführt und die Teilnehmenden erhielten einen Gutschein im Wert von zehn Euro als Aufwandsentschädigung und Anreiz zur Teilnahme. Um auch Familien befragen zu können, deren Muttersprache nicht Deutsch war und die sich das Interview in deutscher Sprache nicht zutrauten, wurden auch Interviews in anderen Sprachen (überwiegend Russisch und Türkisch, sowie vereinzelte Interviews in weiteren Sprachen) ermöglicht.

Die im Rahmen der quantitativen Erhebung erfassten Daten beruhten auf Selbstauskünften der Befragten. In den für die Interviews entworfenen standardisierten Fragebögen wurden sowohl etablierte Skalen als auch eigens konstruierte Items verwendet. Die Items waren überwiegend geschlossen formuliert und wurden durch einige offene Fragen ergänzt. Es wurden Informationen zu den folgenden Themenbereichen erhoben (detailliertere Auskünfte zu den Items und Skalen sind Müller et al., 2015, zu entnehmen):

- Angebotsnutzung
- Fragen zur Fachkraft
- Häusliche Anregungsstrukturen
- Erziehung
- Elterliches Stresserleben
- Entwicklungsstand des Kindes
- Unterstützung
- Bildung
- Außerhäusliche Betreuung, Schule
- Soziodemographie

Auch in Modul 3 wurden die quantitativen Erhebungen durch qualitative Daten aus 41 halbstrukturierten Tiefeninterviews ergänzt (vgl. Abbildung 4). Diese hatten den Zweck, die Situation der Eltern näher zu beleuchten und einen genaueren Einblick in die Zusammenarbeit zwischen Eltern und Elternbegleitern zu geben. Es wurden daher unter anderem vertiefte Informationen zu Nutzungsbarrieren, Belastungen

und Hilfebedarfen sowie zum Kontakt zur Fachkraft erfragt. Zehn Interviews wurden auf Russisch geführt, die restlichen Gespräche fanden in deutscher Sprache statt. Es nahmen insgesamt 40 Frauen und ein Mann an den durchschnittlich eine halbe Stunde dauernden telefonischen Interviews teil und erhielten dafür einen Gutschein im Wert von zehn Euro als Aufwandsentschädigung und Teilnahmeanreiz. Die Interviews wurden transkribiert und inhaltsanalytisch ausgewertet, wobei durch den Einsatz zweier unabhängiger Kodierer und durch die Berechnung der Interraterübereinstimmungen eine objektive und reliable Auswertung verfolgt wurde.

4.1.4 Zentrale Ergebnisse

Abschließend werden die wesentlichen Ergebnisse der Evaluation kurz zusammengefasst. Die Darstellung wird bewusst knapp gehalten, da die Evaluationsergebnisse an anderer Stelle (Beck et al., 2016; Müller et al., 2015) im Detail aufgeführt sind und dort nachgelesen werden können[6]. Zudem sind die Ergebnisse der Evaluation für vorliegende Arbeit weniger zentral als die Rahmenbedingungen und die Methodik.

4.1.4.1 Zentrale Ergebnisse aus Modul 1

Hinsichtlich der Inanspruchnahme der Qualifizierung zeigte sich, dass eine sehr große Nachfrage bestand. Das für den Zeitraum 2011 bis 2014 anvisierte Ziel von 4 000 Elternbegleitern wurde weit übertroffen: Bis Ende des Jahres 2014 hatten bundesweit rund 5 500 Fachkräfte die Qualifizierung erfolgreich durchlaufen und fast 2 000 weitere Fachkräfte bekundeten ihr Interesse, die Qualifizierung zu beginnen. An der Qualifizierung nahmen vor allem Frauen teil; nur knapp 4 Prozent der Befragten waren männlich. Die Altersspanne der Teilnehmenden bewegte sich zwischen 22 und 65 Jahren, wobei insbesondere Fachkräfte zwischen 35 und 50 Jahren die Qualifizierung besuchten. Überwiegend waren die Teilnehmenden der Qualifizierung als Angestellte beschäftigt, aber auch Selbstständige sowie pädagogische Fachkräfte

[6] Es wird hiermit ausdrücklich darauf hingewiesen, dass die in den Abschnitten 4.1.4.1 bis 4.1.4.3 dargestellten Informationen weitgehend wörtlich dem Abschlussbericht der Evaluation (Müller et al., 2015) sowie dem Nachbericht zu Modul 3 (Beck, Schwaß & Stemmler, 2016), jeweils in Absprache mit den Autoren dieser Werke, entnommen sind.

auf Honorarbasis nahmen an der Qualifizierung teil. Mehrheitlich arbeiteten die Teilnehmenden in Voll- oder Teilzeit, nur sehr wenige hatten ein geringfügiges Beschäftigungsverhältnis. Die Qualifizierung besuchten insbesondere ausgebildete Erzieher, aber auch Studienabsolventen waren vertreten. Die befragten Elternbegleiter waren in unterschiedlichen Einrichtungen tätig. Besonders groß war die Nachfrage nach der Qualifizierung bei pädagogischen Fachkräften aus Kindertagesstätten (51 %). Von diesen kam die Hälfte aus Kindertagesstätten, die sich aktuell in Richtung eines Familienzentrums, Eltern-Kind-Zentrums o.Ä. weiterentwickelten. Fachkräfte aus schon bestehenden Familienzentren (17 %) waren häufiger vertreten als Fachkräfte aus Einrichtungen der Familienbildung (8 %). Hauptmotiv für die Teilnahme an der Qualifizierung war es, den Zugang zu Familien bzw. die Ansprache von Eltern zu verbessern. Außerdem erhofften sich viele Fachkräfte, durch die Qualifizierung mehr Wissen über gute und effektive Zusammenarbeit mit den Eltern zum Wohle der Kinder zu erlangen. Es war ihnen ein Anliegen, die Eltern in Gesprächen und Beratungssituationen stärker darin zu sensibilisieren, wie wichtig die Familie für die Bildung der Kinder ist.

Im Rahmen der dreiwöchigen Präsenzphasen während der Weiterbildung wurde eine Vielzahl von Themen behandelt, die von Formen der Zusammenarbeit mit Eltern (z. B. Wertschätzung und Achtsamkeit in der Zusammenarbeit mit Eltern, Kommunikationsmodelle, Beratungsformen und -techniken, Konzepte aufsuchender, niedrigschwelliger Angebote für Eltern) über Inhalte dieser Zusammenarbeit (z. B. Bindung, frühkindliche Entwicklung und Sprache, Bildungsgelegenheiten im Familienalltag, Bildungsverläufe und Bildungsübergänge) bis hin zu selbstbezogenen Kompetenzen (z. B. interkulturelle Kompetenz, Selbstreflexion der eigenen Rolle als Elternbegleiter, Selbstdarstellung und Öffentlichkeitsarbeit) reichten. Die Inhalte der Curricula wurden mithilfe breit gefächerter didaktischer Methoden umgesetzt, wobei vor allem Kleingruppenarbeit und Gruppenarbeit im Plenum, kombiniert mit Vorträgen, im Vordergrund standen. Die verschiedenen Themen wurden unterschiedlich intensiv behandelt, wobei die Varianz in der Tiefe der Themenbehandlung vermutlich bedingt war durch die Orientierung der Kursleiter an Wünschen, Interessen und Vorkenntnissen der Teilnehmenden. Gleichzeitig zeigte sich, dass intensiv behandelte Themen später besser in den Arbeitsalltag integriert werden konnten als weniger intensiv behandelte Inhalte.

Insgesamt waren die Elternbegleiter sehr zufrieden mit der Qualifizierung. Sie fühlten sich zumeist sehr wohl und empfanden die Lernziele als überwiegend klar. Auch mit der Kursleitung sowie dem ausgehändigten Material waren sie zum Großteil sehr zufrieden. Positiv erwähnt wurde, dass die Kursleitung immer zwei Personen umfasste, die sich gegenseitig gut ergänzten. Zudem schätzten die Teilnehmenden die gut gelungene Verknüpfung von Theorie und Praxis und lobten den großen Anteil an Übungen und Möglichkeiten, das neu Gelernte in den Kursen ausprobieren zu können. Fast drei Viertel der Befragten, insbesondere Fachkräfte mit Leitungsposition, hatten vor, das Gelernte an Kollegen weiterzugeben. Die Qualifizierung wies aus Sicht der Fachkräfte eine hohe Praxisrelevanz für die Zusammenarbeit mit Eltern und die Gestaltung von Angeboten der Familienbildung auf. Vier von fünf der befragten Elternbegleiter gaben an, dass sie viele nützliche und unmittelbar umsetzbare Anregungen für ihren Arbeitsalltag erhalten hätten. Insbesondere die wertschätzende Haltung gegenüber Eltern, das vermittelte methodische Repertoire und das Wissen über Bildungsverläufe und Bildungsübergänge wurden als wertvolles „Rüstzeug“ betrachtet. Dennoch vermisste ein gutes Fünftel der Teilnehmenden eine stärkere theoretische Fundierung der Qualifizierung, und ein knappes Drittel wünschte sich, dass einzelne Inhalte mit mehr Tiefe behandelt würden. Dies betraf insbesondere Kenntnisse zu Kommunikationsmodellen, Beratungsformen und -techniken sowie zu niedrigschwelliger Zusammenarbeit mit den Eltern.

Bereits vor Beginn der Qualifizierung arbeiteten die Elternbegleiter nach eigenen Angaben häufiger mit Familien mit Migrationshintergrund, bildungsfernen Familien und Familien in Einkommensarmut zusammen, als dies bei Fachkräften der Kontrollgruppe der Fall war. Nach der Qualifizierung erreichten die Fachkräfte zwar nicht mehr Eltern als vorher, aber die Zusammenarbeit mit den angesprochenen Eltern gelang ihnen nach eigener Einschätzung insgesamt besser. Im Zuge der Qualifizierung konnten Form und Inhalt der Zusammenarbeit mit Eltern stärker in die angestrebte Richtung profiliert werden. So steigerten die Qualifizierungsteilnehmenden bis zur Follow-Up-Erhebung ihre Angebote der Familienbildung und der Bildungsbegleitung, während sich in der Kontrollgruppe keine Veränderungen zeigten. Auch die Häufigkeit, mit der Fragen der Bildung und des Lernens, der allgemeinen Lebensbewältigung und der Sprachförderung in Elterngesprächen zum Thema gemacht wurden, nahm unter den

Teilnehmenden der Qualifizierung zunächst stärker zu als in der Kontrollgruppe, fiel allerdings im Zeitverlauf weitgehend wieder auf das Ausgangsniveau zurück.

Orientiert an den Inhalten der Curricula wurden seitens des Evaluationsteams acht zentrale Kenntnisbereiche identifiziert, in denen Veränderungen im Zeitverlauf untersucht wurden. Die Teilnehmenden begannen die Qualifizierung meist schon mit guten Kenntnissen bezüglich bildungsrelevanter Themen, informeller, niedrigschwelliger und wertschätzender Elternarbeit und Möglichkeiten der Vernetzung im Sozialraum, hatten aber nur in einem Bereich einen Startvorteil gegenüber der Kontrollgruppe (bei ihren Kenntnissen von Konzepten aufsuchender, informeller und niedrigschwelliger Zusammenarbeit mit Eltern). Die Kenntnisse konnten im Zuge der Qualifizierung in allen acht Bereichen noch weiter ausgebaut werden und stiegen auch unter Kontrolle zahlreicher Randbedingungen stärker als in der Kontrollgruppe. Auch einige Monate nach der Qualifizierung waren weiterhin durchweg bessere Kenntnisse vorhanden als vor der Qualifizierung. Ausschlaggebend für die Kenntnisse nach der Qualifizierung war nicht nur das Interesse der Teilnehmenden an den jeweiligen Themen, sondern auch die Intensität, mit der die einzelnen Themenbereiche behandelt worden waren. Angaben aus den qualitativen Interviews bestätigten, dass die Elternbegleiter die in der Qualifizierung vermittelten Kenntnisse über Kommunikationsmodelle sowie Beratungsformen und -techniken als sehr gewinnbringend bewerteten und ihre Gespräche mit Eltern nun weniger intuitiv, sondern professioneller führten.

Zugewinne zeigten sich auch in den fachlichen Kompetenzen. Kompetenzen beinhalten neben notwendigen Kenntnissen insbesondere die Fähigkeit, das Gelernte in unterschiedlichen Situationen in der Praxis umzusetzen. Es wurden Indikatoren für die folgenden Kompetenzbereiche gebildet: (1) Gesprächskompetenz: Anwendung verschiedener Gesprächsstrategien sowie die Fähigkeit, auch in emotional schwierigen Situationen gelassen zu bleiben; (2) berufliche Objektivität: Fähigkeit der Fachkräfte, sich und ihre subjektiven Empfindungen in der Zusammenarbeit mit Eltern zurückzunehmen bzw. ausklammern zu können; (3) zielgruppenspezifische Empathie: Fähigkeit, sich gut in Personen mit anderem kulturellen Hintergrund und Eltern mit schwierigen Lebenslagen hineinversetzen zu können; (4) berufliche Selbstwirksamkeit: Gefühl der Sicherheit in der Arbeit

als Elternbegleiter und (5) konstruktive Zusammenarbeit mit Eltern: Fähigkeit der Fachkräfte, gemeinsam mit den Eltern Lösungen zu erarbeiten. In vier dieser fünf Bereiche schätzten sich die Teilnehmenden nach der Qualifizierung signifikant kompetenter ein als zuvor und hatten hierbei einen höheren Kompetenzzuwachs als die Kontrollgruppe. Lediglich hinsichtlich der beruflichen Objektivität waren keine Veränderungen zu verzeichnen. Insgesamt hatte die Qualifizierung einen schwächeren Effekt auf die Kompetenzen der Fachkräfte als auf deren Kenntnisse. Obwohl sich in der Folgebefragung (Follow-Up) in drei von fünf Bereichen wieder ein Rückgang der Kompetenzen abzeichnete, schien der Kompetenzzuwachs auch mehrere Monate nach der Qualifizierung weiterhin Bestand zu haben.

Laut Auskunft der Fachkräfte in der Follow-Up-Befragung konnte vor allem das erworbene Wissen um Wertschätzung und Achtsamkeit in der Zusammenarbeit mit Eltern sehr gut umgesetzt und in die eigene Arbeit integriert werden, gefolgt von Kenntnissen zu Kommunikationsmodellen, Beratungsformen und -techniken. Weniger leicht fiel die Umsetzung in den Bereichen Bildungs- und Erziehungsvorstellungen in unterschiedlichen sozialen Milieus und Möglichkeiten der Vernetzung im Sozialraum. Das Schlusslicht lieferten die Bildungssysteme in den Bundesländern. In einzelnen Themenbereichen hatten Führungskräfte Vorteile bei der Umsetzung (z. B. Kooperation und Vernetzung). Auch schienen Kindertagesstätten und Familienzentren einen günstigeren Kontext für die Umsetzung zu bieten.

In diesem Zusammenhang wurde auch untersucht, welche Bedingungen die Arbeit als Elternbegleiter erleichterten. Hierzu nannten die Fachkräfte vor allem das positive Feedback der Eltern, ihr persönliches Interesse und die Unterstützung durch Arbeitskollegen. Als wichtigste Erschwernisse, die einer Integration der Elternbegleitung in die täglichen Arbeitsabläufe entgegenstanden, wurden das geringe Zeitkontingent, die fehlende Finanzierung und eine hohe Arbeitsbelastung, häufig auch verursacht durch Personalmangel und einen hohen Krankenstand, genannt. Insbesondere die qualitativen Interviews der Fachkräfte zeigten, dass die in den Fortbildungen erworbenen Kenntnisse und Fähigkeiten vor allem dann in der Praxis wirksam wurden, wenn die Elternbegleiter in die Team- und Organisationsentwicklung ihrer Einrichtung eingebunden waren und alle Beteiligten den Gedanken der Elternbegleitung mittrugen. Als besonders günstig erwies es sich, wenn die Einrichtung bzw. der Träger bereits ein Konzept zur

Zusammenarbeit mit Eltern hatte und dieses konsequent, aber auch lernoffen umsetzte (z. B. Öffnung einer Familienbildungsstätte in den Sozialraum oder Weiterentwicklung einer Kita zum Familienzentrum). Nicht zuletzt gelang der Transfer auch besser, wenn das gesamte Team über die Ziele der Elternbegleitung informiert war und sich mit diesen identifizierte, statt die Elternbegleitung als Konkurrenz oder Einmischung zu sehen.

4.1.4.2 Zentrale Ergebnisse aus Modul 2

Bei der Mehrheit der *Elternbegleitung Plus*-Modellstandorte handelte es sich um Einrichtungen der Familienbildung (52 %), gefolgt von Eltern-Kind-Zentren bzw. Familienzentren (22 %), Kindertagesstätten, die sich zu Familienzentren entwickelten (8 %) und herkömmlichen Kindertagesstätten (4 %). Weitere 15 Prozent waren andere Einrichtungen. Überwiegend waren die Einrichtungen in einer mittelgroßen Stadt (43 %) oder Großstadt bzw. Metropole (37 %) angesiedelt, während nur wenige Einrichtungen in Kleinstädten (17 %) oder im ländlichen Raum (2 %) lagen. Das räumliche Einzugsgebiet der geförderten Einrichtungen beschränkte sich vielfach nicht nur auf die jeweilige Stadt oder Gemeinde, sondern ging darüber hinaus auch in benachbarte Gemeinden, das Umland oder andere Landkreise und Städte.

Ein Großteil der Einrichtungen schätzte den Anteil der Zielgruppen von Elternbegleitung (Familien mit Migrationshintergrund, bildungsbenachteiligte Familien und Familien in Einkommensarmut) in ihrem Einzugsgebiet als hoch oder sehr hoch ein. Diesen Zielgruppen wurde im Rahmen der Fokusgruppen seitens der Teilnehmer zwar vielfach ein erhöhter Bedarf an Förderung und Begleitung zugeschrieben, sei es aufgrund mangelnder Sprachkenntnisse oder aufgrund nur schwach ausgeprägter Erziehungskompetenzen. Gleichzeitig wurde aber auch auf Zugangshürden aufgrund mangelnden Vertrauens seitens der Eltern verwiesen. Vor allem bildungsferne Eltern ohne Migrationshintergrund wurden als schwer erreichbar gesehen.

Rund zwei Drittel der Modelleinrichtungen finanzierten zwischen vier und acht Angebote über die Förderung *Elternbegleitung Plus*. Am häufigsten wurden offene Treffs und Gesprächskreise angeboten (von 69 % der Einrichtungen), gefolgt von Einzelgesprächen (61 %) und Einzelberatung (56 %) für Eltern. Rund die Hälfte der Einrichtungen bot jeweils einzelne Vorträge oder Informationsabende an, wobei auch

feste Kurse (48 %) und Eltern-Kind-Gruppen (44 %) ähnlich häufig vertreten waren. Hausbesuche (33 %) und Begleitung zu Ämtern, Schulen etc. (32 %) wurden demgegenüber seltener angeboten. In der Regel wurden mehrere Angebotsformen kombiniert, um Eltern niedrigschwellig zu erreichen.

Zugänge zur Zielgruppe wurden überwiegend über die eigene Einrichtung gesucht, indem dort auf dieses Angebot hingewiesen wurde (75 %) oder potenzielle Interessierte persönlich angesprochen wurden (74 %). Bei 61 Prozent der Einrichtungen erfolgte der Zugang zu Eltern immer oder häufig mittels Mund-zu-Mund-Propaganda unter den Eltern. Bei 59 Prozent der Einrichtungen wurden Interessierte persönlich außerhalb der Einrichtung angesprochen, und in 37 Prozent der Fälle erfolgte die Akquise über Angebote in anderen Einrichtungen. Andere Einrichtungen waren aber nicht nur in die Teilnehmerakquise eingebunden, sondern auch in die Durchführung der Angebote, die sogar vergleichbar oft in Räumlichkeiten von Kooperationspartnern stattfand wie in der eigenen Einrichtung. Nur selten (9 % bis 12 %) wurden öffentliche Plätze für die Aktivitäten genutzt.

Die Fokusgruppen legten nahe, dass im Rahmen von *Elternbegleitung Plus* insbesondere die persönliche Ansprache der Eltern in den Vordergrund rückte. Dabei ging es den Fachkräften nicht nur um eine erfolgreiche Teilnahme der Eltern an einem bestimmten Angebot, sondern mehr noch um den langfristigen Aufbau einer vertrauensvollen Beziehung, die als Vorbedingung für die erfolgreiche Vermittlung von Eltern der Zielgruppen in geeignete Angebote (z. B. Elterncafé, Elternkurse) gesehen wurde. Bei der persönlichen Ansprache wurde die wertschätzende und partnerschaftliche Haltung gegenüber den Eltern als besonders wichtig erachtet. Werbung in Form von Flyern, Plakaten oder Aushängen wurde hinsichtlich der Erfolgschancen demgegenüber eher skeptisch beurteilt. Um die regelmäßige Teilnahme von neu gewonnenen Eltern zu sichern, wurde auf verschiedene Strategien zurückgegriffen (z. B. telefonische Erinnerung vor Beginn des Angebots, erneute Kontaktierung der Eltern bei Nicht-Einhaltung eines Termins). Die Teilnahme an regelmäßigen Angeboten wurde attraktiver gemacht, indem auf die Zusammensetzung der Gruppen sowie auf die Inhalte und Form der Kurse geachtet wurde. Auch durch die Abstimmung und zeitliche Verkettung von Angeboten wurde versucht, Eltern längerfristig an die Einrichtungen zu binden.

Die Kooperationen der Modellstandorte zielten vor allem darauf ab, neue Zugangswege zu entwickeln und neue Zielgruppen zu erreichen. Am häufigsten bestand eine Kooperation mit Kindertagesstätten (69 %), gefolgt von Kooperationen mit Erziehungs- und Familienberatungsstellen (58 %). Rund die Hälfte der befragten Standorte kooperierte mit Grundschulen (52 %), Familienzentren (47 %) und Jugendämtern (46 %). Gut ein Drittel der Befragten arbeitete im Rahmen der Förderung auch mit Einrichtungen der Sozialpädagogischen Familienhilfe zusammen (39 %), während jeweils circa ein Viertel mit Mehrgenerationenhäusern (26 %), Migrationsdiensten und Kinderärzten (jeweils 25 %) kooperierte. Insgesamt ergab sich eine große Bandbreite von Kooperationspartnern, die auch keineswegs auf Institutionen unter gleicher Trägerschaft begrenzt waren. Beim Aufbau des Kooperationsnetzwerkes im Rahmen von *Elternbegleitung Plus* intensivierten die Modellstandorte vor allem die Zusammenarbeit mit bereits bekannten Partnern. Aber auch 62 neue Kooperationen entstanden, die sich auf zahlreiche Kooperationspartner erstreckten. Auch unter diesen neuen Kooperationspartnern hatten Kindertagesstätten eine Spitzenposition.

Im Rahmen der Fokusgruppen wurde diskutiert, wodurch sich Kooperationen auszeichnen. Hierbei wurde Kooperation als institutionelle Zusammenarbeit in Bezug auf ein konkretes Projekt oder Ziel definiert. Als wichtig wurde eine mündliche oder schriftliche Vereinbarung erachtet, wobei auch herausgestellt wurde, dass formelle Abmachungen für sich allein keine erfolgreiche Kooperation ermöglichen, sondern das Engagement der Leitungskräfte sowie der Mitarbeiter der verschiedenen Einrichtungen entscheidend ist. Im Unterschied zur Kooperation wurde das Netzwerk als unverbindlicher, informeller Kontakt zwischen verschiedenen Partnern dargestellt, der keinem konkreten Ziel dient. Für den Aufbau und Erhalt einer kontinuierlichen und verbindlichen Beziehung zwischen den Partnern wurden in den Fokusgruppen neben gemeinsamen Zielen und Gegenseitigkeit als weitere wichtige Faktoren genannt: persönliche Kontakte im Vorfeld, die eine Vertrauensbasis schaffen, räumliche Nähe, regelmäßiger Austausch, zeitliche, finanzielle und stabile personelle Ressourcen und nicht zuletzt die Unterstützung durch die Einrichtungsleitung. Letztere wurde insbesondere dann als wichtig angesehen, wenn nicht schon tragfähige persönliche Kontakte bestehen und die Kooperation nahelegen.

Auch die Rekonstruktion der Kooperationsnetzwerke der Modellstandorte, die an den Fokusgruppen teilnahmen, bestätigte die zentrale Rolle der Kindertagesstätten als Partnereinrichtungen von *Elternbegleitung Plus*, da sie den Zugang zu den Zielgruppen des Programms über persönliche Ansprache erleichterten und den Zugang zu Grundschulen und somit die Entwicklung von Strukturen der Bildungsbegleitung unterstützten. Häufig wurden seitens der Familienbildungsstätten Elternbegleiter als externe Fachkräfte in den Einrichtungen der Kooperationspartner, vor allem Kindertagesstätten, eingesetzt. Diese Arbeit war erfolgreich, wenn es gelang, eine vertrauensvolle Kooperationsbasis zum Personal der Einrichtung aufzubauen, während mangelndes Vertrauen der Mitarbeiter gegenüber der externen Fachkraft eine Zusammenarbeit stark beeinträchtigte und eine Vermittlung von Eltern unwahrscheinlich machte.

Jenseits des Vorschulbereichs, in dem der Bedarf an Elternbegleitung bei der Vorbereitung auf den Schuleintritt besonders hoch war, gestaltete sich die Kooperation mit Schulen jedoch schwierig, da trotz der hohen Relevanz, die der Bildungsbegleitung in Schulen zugeschrieben wurde, der Zugang zu Schulen kaum gelang. Als Gründe hierfür wurden eine geringere Akzeptanz der Zusammenarbeit mit Eltern im Schulkontext, aber auch das Fehlen einer Kommunikation auf Augenhöhe gesehen.

Schon bei der Erstbefragung waren an den Modellstandorten durchschnittlich ein bis zwei qualifizierte Elternbegleiter tätig und weitere Qualifizierungen des Personals für Elternbegleitung waren vorgesehen. In das breit gefächerte Angebot, das innerhalb der Kooperationsnetzwerke für Eltern bereitgestellt wurde, waren Elternbegleiter je nach Zuschnitt der Angebote in unterschiedlichem Maße eingebunden. Vorwiegend waren sie in offenen Treffs und Gesprächskreisen (71 %), Einzelgesprächen mit Eltern (69 %) und Eltern- bzw. Informationsabenden (62 %) tätig, hielten einzelne Vorträge (46 %) oder führten Eltern-Kind-Gruppen durch (45 %). Unter den Formen aufsuchender Arbeit übernahm rund jede dritte qualifizierte Fachkraft die Begleitung von Eltern zu Ämtern, Schulen oder Beratungseinrichtungen (34 %), jede fünfte war in mobiler Bildungs- oder Sozialarbeit vor Ort tätig (19 %) und mindestens jede vierte übernahm Hausbesuche (davon 7 % Hausbesuche im Rahmen von Programmen, 25 % andere Hausbesuche). Dies legt nahe, dass die Elternbegleiter über die Modellstandorte in stärkerem Maße aufsuchende Zugänge umsetzten

als qualifizierte Fachkräfte per se. Vielfach übernahmen die Elternbegleiter aber auch Aufgaben jenseits der Zusammenarbeit mit Eltern. Dies betraf insbesondere die Supervision und kollegiale Beratung (31 %) sowie die Fort- und Weiterbildung von anderen Fachkräften (28 %), teilweise auch die Vermittlung von weiterführenden Hilfen (51 %) oder Kinderbetreuung (13 %). Die qualitative Rekonstruktion einzelner Kooperationsnetzwerke lässt darauf schließen, dass ein Teil der Fachkräfte keinen direkten Kontakt mit Eltern hatte, sondern vorrangig mit der Organisation von Angeboten und der Koordination der Fachkräfte betraut war. Insbesondere galt dies für Elternbegleiter in Leitungspositionen. Insgesamt schätzten Elternbegleiter in den Fokusgruppen die Einbindung in den *Elternbegleitung Plus*-Kooperationsnetzwerken sehr, auch weil dies den Einsatz der qualifizierten Fachkräfte in Partnereinrichtungen erleichterte. Auch die Vermittlung von Eltern von einer Einrichtung zur anderen wurde durch die Netzwerke deutlich unterstützt. In dieser Hinsicht erwies sich die Förderung institutioneller Vernetzung als gute Ergänzung zur Qualifizierung der Elternbegleiter.

4.1.4.3 Zentrale Ergebnisse aus Modul 3

Bei den befragten Elternteilen handelte es sich überwiegend um Frauen, nur 5 Prozent der Interviewpartner waren männlich. Die Befragten der Interventionsgruppe waren im Durchschnitt 34 Jahre alt, die Eltern der Kontrollgruppe hatten ein durchschnittliches Alter von 35 Jahren. In beiden Gruppen lebten im Haushalt durchschnittlich zwei Kinder. In 89 Prozent der Fälle in der Interventionsgruppe wurde das Interview auf Deutsch geführt, in der Kontrollgruppe lag dieser Anteil bei 98 Prozent. Hinsichtlich der Zielgruppen des Bundesprogramms zeigte sich, dass diese in der Interventionsgruppe durchweg häufiger vertreten waren als in der Kontrollgruppe. So hatten 39 Prozent der Eltern in der Interventionsgruppe und 24 Prozent der Kontrollgruppeneltern einen Migrationshintergrund entsprechend der Definition des Statistischen Bundesamtes. Während ein Fünftel der Befragten der Interventionsgruppe maximal den Qualifizierenden Hauptschulabschluss besaß und somit als niedrigqualifiziert galt, war dies bei lediglich 5 Prozent der befragten Eltern der Kontrollgruppe der Fall. Und auch der Anteil an armutsgefährdeten Familien war in der Interventionsgruppe (26 %) deutlich höher als in der Kontrollgruppe (11 %).

Die Elternbefragung machte deutlich, dass Elternbegleitung nicht nach einem einheitlichen Muster stattfand, sondern in der Praxis verschieden ausgestaltet wurde und sich in unterschiedlichen Formen der Zusammenarbeit mit Eltern manifestierte. So existierten neben der langjährigen Zusammenarbeit mit Eltern auch gerade erst entstandene Kontakte und auch die Häufigkeit der Treffen mit der Fachkraft variierte. Meist kannten sich Elternbegleiter und Eltern über eine Einrichtung der Kinderbetreuung. Im Rahmen der Elternbegleitung wurden häufig Informations- oder Beratungsgespräche geführt und Veranstaltungen wie Elterncafés, Eltern-Kind-Gruppen oder informative Vorträge angeboten, die durch informelle Tür-und-Angel-Gespräche ergänzt wurden. Hausbesuche, Ausflüge oder eine Begleitung zu Ämtern wurden ebenfalls, wenn auch eher selten, von den Eltern genutzt. Die Zusammenarbeit fand überwiegend im Büro der Fachkraft oder in der Einrichtung statt und nur in Ausnahmefällen bei den Eltern zu Hause. Es wurden im Rahmen der Elternbegleitung verschiedene Themen behandelt. Dazu gehörten in der Regel Fragen zur Entwicklung und Erziehung des Kindes, zu Lerngelegenheiten im Alltag oder auch zur Einschulung. Ein Teil der Elternbegleiter übernahm zudem die Rolle einer Vermittlungsinstanz, indem den Eltern Angebote oder Ansprechpersonen empfohlen wurden. Diese Empfehlungen wiesen eine überwiegend hohe Passgenauigkeit zu den Bedürfnissen und Situationen der Eltern auf; dies wurde u.a. dadurch bestätigt, dass die Eltern eine sehr hohe Zufriedenheitsbewertung in Bezug auf die Betreuung durch die Elternbegleiter abgaben. So waren nahezu alle Eltern (99 %) zufrieden mit ihrem Elternbegleiter, die überwiegende Mehrheit (79 %) war sogar sehr zufrieden. Zudem gaben 94 Prozent der Eltern an, ihren Elternbegleiter weiterempfehlen zu wollen. An der Arbeit der Fachkräfte schätzten die Eltern vor allem deren Wissen und Engagement sowie gut verständliche Erklärungen und Ratschläge. Veranstaltungen wie z. B. Elterncafés, die von Elternbegleitern angeboten wurden, stießen bei den teilnehmenden Eltern auf Lob. Als besonders hilfreich empfanden Eltern dabei den informativen Austausch mit ihrem Elternbegleiter und mit anderen Eltern in einer angenehmen Gruppenatmosphäre. Für Eltern mit Migrationshintergrund, deren Familiensprache nicht Deutsch war, war es zudem wichtig, dass der Elternbegleiter ihre Muttersprache sprach. Vertrauen und Wertschätzung prägten die Beziehung der Eltern zu den Elternbegleitern: Zweiundachtzig Prozent der Eltern gaben an, viel Vertrauen zu ihnen zu

haben. Die Eltern fühlten sich mehrheitlich gut angenommen und erfuhren durch die Elternbegleitung große Wertschätzung. Besonders gut gelang den Fachkräften der Aufbau einer vertrauensvollen Beziehung bei der Subgruppe der armutsgefährdeten Eltern.

Der Vergleich mit der Kontrollgruppe machte deutlich, dass hier ebenso vertrauensvolle Beziehungen zur Fachkraft bestanden wie in der Interventionsgruppe. So ergaben sich für die Zufriedenheit mit und das Vertrauen zu der Fachkraft ebenso wie für die empfundene Wertschätzung und die Bereitschaft, die Fachkraft weiterzuempfehlen, keine Gruppenunterschiede. Positive Effekte der Elternbegleitung wurden besonders in der Nutzung der Fachkräfte als gezielte Ansprechpersonen der Eltern sichtbar. So nahmen Eltern der Interventionsgruppe insgesamt mehr verschiedene Gelegenheiten zur Zusammenarbeit mit ihrem Elternbegleiter wahr. Auch wurde die Elternbegleitung häufiger dazu genutzt, bildungsbezogene Themen zum eigenen Kind anzusprechen. Allerdings waren die Kinder dieser Eltern im Mittel etwas älter als die Kinder in der Kontrollgruppe, wodurch bildungsbezogene Themen zum Befragungszeitpunkt vermutlich eine höhere praktische Relevanz für die Interventionsgruppe hatten. Weitere positive Effekte der Elternbegleitung zeigten sich darin, dass die Eltern hinsichtlich der Vermittlung von zusätzlichen Angeboten oder Ansprechpersonen weniger Bedarf hatten als die Eltern der Kontrollgruppe. Das ist möglicherweise dadurch zu erklären, dass die Elternbegleitung die Bedürfnisse der Eltern umfangreicher abzudecken vermochte.

Im Hinblick auf Veränderungen in der Nutzung von Bildungsangeboten sowie der Kenntnis von Ansprechpersonen deutete sich eine Wirkung der Elternbegleitung auf die Wahrnehmung von Angeboten für Kinder in den befragten Familien an: Die Elterngruppe mit der längsten Begleitung durch Elternbegleiter gab an, die meisten Kinderangebote wahrzunehmen. Für Elternangebote zeigte sich dieses Ergebnis jedoch nicht, hier nahm die Kontrollgruppe ohne Elternbegleitung die meisten Angebote wahr. Auch die von den Eltern beschriebenen Nutzungsbarrieren unterschieden sich nicht zwischen den Gruppen. Wurde nicht an Angeboten teilgenommen, bestanden die Nutzungsbarrieren vor allem in ungünstigen organisatorischen Bedingungen wie fehlender Zeit der Eltern und Schwierigkeiten bei der Kinderbetreuung, manchmal aber auch im fehlenden Interesse an den speziellen Inhalten der Angebote. Mit der Auswahl an Angeboten zeigten

sich beide Untersuchungsgruppen gleichermaßen zufrieden, jedoch brachten Eltern, die sich in Elternbegleitung befanden, tendenziell weniger Verbesserungsvorschläge in Bezug auf die Rahmenbedingungen von Bildungsangeboten hervor. Darüber hinaus wurde deutlich, dass Eltern mit länger bestehendem Kontakt zu einem Elternbegleiter über eine bessere Kenntnis von Ansprechpersonen bei Bildungsfragen verfügten.

Was den Alltag der Familien angeht, so zeigte die Evaluation, dass Elternbegleitung zwar keinen Einfluss auf den Kenntnisstand der Eltern zum deutschen Schulsystem hatte, jedoch die elterlichen Befürchtungen zum Schuleintritt reduzieren konnte, die sie in Bezug auf die Anforderungen hatten, die der Übergang für ihr Kind darstellte. Auf sich als Eltern bezogene Befürchtungen zum Schuleintritt veränderten sich durch die Elternbegleitung kaum und die elterliche Einschätzung der eigenen Fähigkeit, die Kinder bei der Einschulung zu unterstützen, war in allen Gruppen schon zu Beginn der Untersuchung so hoch ausgeprägt, dass keine signifikante Verbesserung mehr möglich war. Der Vergleich der Eltern in Elternbegleitung mit der Kontrollgruppe zeigte außerdem, dass familiäre Anregungsstrukturen im Haushalt bei Kontaktaufnahme zu einem Elternbegleiter zwar anstiegen, dies jedoch auch ohne Elternbegleitung ebenso deutlich taten. Sowohl die elterliche Belastung als auch die kindliche Entwicklung im sprachlichen, kognitiven, emotionalen und motorischen Bereich waren von der Elternbegleitung unbeeinflusst. Letztlich galt dasselbe für das elterliche erziehungsbezogene Kompetenzerleben, jedoch konnten armutsgefährdete Teilnehmer hier durchaus von der Elternbegleitung profitieren: In dieser Gruppe erbrachte die Elternbegleitung eine Steigerung des Selbstwirksamkeitserlebens der Eltern in der Erziehung ihrer Kinder.

4.2 Bewertung der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“: Die Meta-Evaluation

Im Folgenden wird die Bewertung der Evaluation des Bundesprogrammes „Elternchance ist Kinderchance“ dargelegt. Zunächst wird die Methode dieser Meta-Evaluation erläutert, anschließend erfolgt die eigentliche Bewertung anhand von Evaluationsstandards. Die resultierenden Ergebnisse werden abschließend zusammengefasst und im nachfolgenden Kapitel diskutiert.

4.2.1 Methode der Meta-Evaluation

Die vorliegende Evaluationsbewertung stellt eine interne Meta-Evaluation dar, da die Bewertung durch ein Mitglied des Evaluationsteams vorgenommen wurde. Sie fand nach dem Ende der Evaluationsstudie statt und kann daher als summative bzw. retroaktive Meta-Evaluation eingeordnet werden (vgl. Abschnitt 2.2.4)[7]. Es handelte sich dabei nicht um eine beauftragte Meta-Evaluation, die etwa aufgrund eines Konfliktfalles erforderlich gewesen wäre, sondern um eine Forschungsarbeit mit wissenschaftlichem Erkenntnisinteresse. Die Zielsetzung dieser Meta-Evaluation bestand darin, sich kritisch mit der eigenen Arbeit auseinanderzusetzen und Empfehlungen für künftige Evaluationen ableiten zu können. Zu diesem Zweck wurde die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ systematisch daraufhin geprüft, welche Evaluationsstandards eingehalten worden waren und welche nicht. Sowohl die Einhaltung als auch die Nicht-Einhaltung von Standards sollte Hinweise auf zentrale Herausforderungen im Evaluationsprozess liefern und die Generierung von Empfehlungen ermöglichen (vgl. Kapitel 3).

Die Meta-Evaluatorin war von Mai 2012 bis zum Ende der Evaluationsstudie Mitglied des Evaluationsteams. Sie war an der FAU angestellt und somit mit Modul 3 der Evaluation befasst. Die Evaluation des

[7] Zwar fand ein Teil der Meta-Evaluationsaktivitäten (der erste Bewertungsdurchgang sowie die Erhebung der Stellungnahmen, vgl. Abschnitt 4.2.1) noch während der Laufzeit der Evaluation bzw. während der finalen Arbeiten am Abschlussbericht statt, jedoch hatte die Meta-Evaluation keinen Einfluss auf die Evaluation und verfolgte auch nicht den Zweck, diese zu verbessern. Daher handelte es sich nicht um eine formative, sondern um eine summative Meta-Evaluation.

Bundesprogramms „Elternchance ist Kinderchance“ stellte ihre erste Tätigkeit als Evaluatorin dar, und im Rahmen der vorliegenden Arbeit übernahm sie erstmalig die Rolle einer Meta-Evaluatorin. Da die Meta-Evaluation auch eine Bewertung ihrer eigenen Tätigkeit als Evaluatorin beinhaltete, bestand die Gefahr einer verzerrten Urteilsbildung. Dessen bewusst, traf die Meta-Evaluatorin verschiedene Maßnahmen, um die Subjektivität ihres Urteils zu reduzieren. Diese werden untenstehend erläutert.

Um die Bewertung der Evaluation von „Elternchance ist Kinderchance“ vornehmen zu können, erfolgte zunächst eine intensive Einarbeitung in die Evaluationsstandards durch die Meta-Evaluatorin. Da es sich um ein deutsches Evaluationsprojekt handelte, wurden die deutschen Standards der DeGEval verwendet (DeGEval, 2008b; vgl. Abschnitt 2.2.3). Diese erwiesen sich allerdings als zu wenig konkret gefasst, um daraus Kriterien für die Evaluationsbewertung ableiten zu können. Daher wurde zudem das Handbuch zu den amerikanischen Evaluationsstandards herangezogen (JCSEE, 2006), welches deutlich ausführlichere Erläuterungen enthält und von der DeGEval als Ergänzung zu ihren Standards explizit benannt wird. Aus diesen beiden Werken wurden für jeden der 25 Einzelstandards Bewertungskriterien abgleitet, indem die in den DeGEval-Richtlinien genannten Anforderungen um die des JCSEE-Handbuchs ergänzt wurden. Die Kriterien für jeden Standard sind in Abschnitt 4.2.2 angegeben. Lediglich einzelne in den Standards enthaltene Kriterien wurden nicht einbezogen, da sie für die vorliegende Evaluation nicht passend waren. Dies betraf z. B. solche Hinweise, die sich (aufgrund des Erziehungskontexts der amerikanischen Standards) speziell auf Programme in Schulen und dementsprechend auf Schüler, Eltern und Lehrer bezogen und für die Evaluation von „Elternchance ist Kinderchance“ nicht zutrafen. In Abschnitt 4.2.2 erfolgt daher keine vollständige Aufzählung aller Handbuch-Kriterien, vielmehr wird auf die Nennung der nicht zutreffenden Anforderungen aus darstellungsökonomischen Gründen verzichtet. Auf dieses selektive Vorgehen wird unter dem Gesichtspunkt der Nachvollziehbarkeit im Diskussionsteil eingegangen (vgl. Abschnitt 5.3.1).

Für die Bewertung der Evaluation wurden vier Kategorien erstellt:

1. **Standard wurde erfüllt**: Diese Bewertung wurde vergeben, wenn alle Anforderungen, die in den deutschen und amerikanischen Evaluationsrichtlinien zu diesem Standard enthalten waren, von der Evaluation eingehalten worden waren.
2. **Standard wurde überwiegend erfüllt**: Waren lediglich geringfügige Verstöße zu verzeichnen, wurde der Standard als überwiegend erfüllt betrachtet.
3. **Standard wurde teilweise erfüllt**: Diese Beurteilung erfolgte, wenn mittelschwere Verstöße gegen die Standardanforderungen aufgedeckt wurden, die sich jedoch mit den eingehaltenen Bewertungskriterien die Waage hielten.
4. **Standard wurde nicht erfüllt**: Ein Standard wurde dann als nicht erfüllt eingestuft, wenn hinsichtlich der Bewertungskriterien viele Abzüge vorlagen, denen keine oder nur sehr wenige eingehaltene Kriterien gegenüberstanden.

Während die Bewertungskategorien 1 und 4 für Außenstehende plausibel sein dürften, muten die Stufen 2 und 3 vermutlich subjektiv an, da keine Operationalisierung von „geringfügigen“ und „mittelschweren“ Verstößen erfolgt. Zu beachten ist, dass bei der vorliegenden Meta-Evaluation keine numerische Bewertung im Sinne eines Summenscores vorgenommen werden sollte. Vielmehr sollten die Stärken und Schwächen der Evaluation möglichst vollständig abgebildet und nachvollziehbar beschrieben werden, um Empfehlungen zu ermöglichen. Das Urteil selbst, also wie viele Standards von der Evaluation eingehalten worden waren, war dabei zweitrangig. Dennoch wurde versucht, die Subjektivität des Urteils einzudämmen. Eine der Maßnahmen, die zu diesem Zweck ergriffen wurden, bestand darin, dass die konservativere Bewertungsstufe gewählt wurde, wenn die Daten nicht eindeutig für eine bestimmte Kategorie sprachen. Legten die Daten z. B. eine Bewertung eines Standards als entweder *überwiegend erfüllt* oder *teilweise erfüllt* nahe, so wurde in solchen Fällen die konservativere Bewertung *teilweise erfüllt* vergeben.

Es sei darauf hingewiesen, dass bei der Unterscheidung zwischen *überwiegend*, *teilweise* und *nicht erfüllten* Standards sowohl die Anzahl als auch die Schwere der Verstöße einbezogen wurde. Die Schwere bezog sich dabei auf die unterschiedliche Gewichtung von Bewertungs-

kriterien, von denen manche als sehr wichtig für das Gelingen einer Evaluation gesehen wurden, andere dagegen als etwas weniger zentral. Dementsprechend wurde die Nicht-Einhaltung sehr bedeutsamer Faktoren als gravierender bewertet als die Nicht-Einhaltung von Kriterien, die für die Evaluation von untergeordneter Relevanz waren. Während sich die Anzahl der Verstöße leicht in numerischer Form (z. B. anhand einer Checkliste) abbilden ließe, erschien selbiges für die Schwere nicht pauschal möglich oder sinnvoll. Die Operationalisierung der Urteile wurde zudem dadurch erschwert, dass bei einigen Standards eine Relativierung an den zur Verfügung stehenden Ressourcen erfolgen und/oder die Einordnung der Evaluation als auftraggeberorientiert berücksichtigt werden musste. Auch aus diesem Grund erschien eine nähere Explikation der vier Kategorien, z. B. im Sinne von „Die Bewertung *überwiegend erfüllt* wurde vergeben, wenn höchstens zwei Verstöße gegen maximal als wenig relevant eingestufte Kriterien erfolgt waren“, als wenig sinnvoll. Dieses Vorgehen bei der Evaluationsbewertung wird, insbesondere hinsichtlich der beiden Aspekte der Subjektivität und der Nachvollziehbarkeit, bei der Methodenkritik im Diskussionsteil (Abschnitt 5.3.1) aufgegriffen.

Nach der Erarbeitung der Bewertungskriterien und -kategorien wurde die Evaluation darauf untersucht, wie viele der Bewertungskriterien sie erfüllt hatte und in welchem Ausmaß der jeweilige Standard demnach als eingehalten beurteilt werden konnte. Für diesen Schritt standen der Meta-Evaluatorin, die fast von Beginn der Evaluation an im Evaluationsteam der FAU tätig gewesen war, das projektbezogene Wissen und die im Verlauf der Evaluation gesammelten Erfahrungen und Daten zur Verfügung. Darüber hinaus orientierte sie sich bei der Bewertung an folgenden Dokumenten und Unterlagen: Auftragsbekanntmachung, Angebotsaufforderung, Auftragsbeschreibung einschließlich Anlagen, Angebot des Evaluationsteams einschließlich Kostenkalkulation, Werkvertrag zwischen Evaluatoren und Auftraggeber einschließlich Änderungsverträgen, Kooperationsvereinbarung und Datenaustauschvereinbarung zwischen DJI und FAU, Projektberichte sowie Protokolle und sonstige Aufzeichnungen aus dem Evaluationsprozess. Auf Grundlage dieser Informationen erfolgte im Sommer 2014 eine erste Bewertung der Evaluation anhand der 25 Evaluationsstandards. Diese Bewertung musste jedoch als vorläufig gelten, da zum einen die Evaluation zu diesem Zeitpunkt noch nicht

abgeschlossen war und somit einige Aspekte, z. B. den Abschlussbericht betreffend, noch nicht endgültig beurteilt werden konnten. Zum anderen stellte sich heraus, dass die Meta-Evaluatorin bei einigen Standards keine Einschätzung vornehmen konnte, weil die ihr zur Verfügung stehenden Informationen für ein abschließendes Urteil nicht ausreichten. Dies war beispielsweise beim Standard N3 (Glaubwürdigkeit und Kompetenz des Evaluators/der Evaluatorin) der Fall, bei dem sie zwar angeben konnte, welche Maßnahmen die Evaluatoren ergriffen hatten, um ihre Glaubwürdigkeit zu erhöhen. Ob diese jedoch tatsächlich zu einer hohen wahrgenommenen Glaubwürdigkeit geführt hatten, darüber konnten nur andere an der Evaluation beteiligte Personen Aufschluss geben.

Aufgrund dieser Erkenntnis holte die Meta-Evaluatorin von Januar bis Mai 2015 Stellungnahmen von Vertretern verschiedener Stakeholder ein. Dazu sammelte sie diejenigen Aspekte der Standards, zu welchen sie Auskünfte benötigte, in einem Fragenkatalog, welchen die Stellungnehmer entweder schriftlich oder mündlich in Form eines Telefoninterviews beantworten konnten. Die Fragen umfassten unter anderem die Beteiligung während des Evaluationsprozesses, Informationen über Ergebnisse, die Nutzung der Ergebnisse und die Zusammenarbeit mit dem Evaluationsteam. Viele Fragen wurden allen Stakeholdern gestellt, einige Informationen hingegen nur bei bestimmten Stellungnehmern erhoben, die über die entsprechenden Einsichten in den Evaluationsprozess verfügten. Der Fragenkatalog einschließlich Angaben dazu, welche Fragen welchen Stellungnehmern gestellt wurden, kann dem Anhang entnommen werden.

Ziel der Stellungnahmen war es, die Perspektiven möglichst vieler Stakeholdergruppen einzubeziehen und bei der Evaluationsbewertung zu berücksichtigen, um zu verlässlicheren, weniger subjektiven Urteilen zu kommen. Zu diesem Zweck wurden Vertreter jedes Stakeholders angesprochen und um eine Stellungnahme gebeten, mit Ausnahme der wissenschaftlichen Impulsgruppe, da diese als Beirat lediglich punktuelle Einsichten in den Evaluationsprozess gehabt hatte und die Stellungnahmen auf Personen abzielten, die direkter an der Evaluation oder am Bundesprogramm beteiligt gewesen waren. Erfreulicherweise wurde die Meta-Evaluation sehr unterstützt und es konnten insgesamt 12 Stellungnahmen eingeholt werden (obwohl dafür keine Aufwandsentschädigung angeboten wurde). Diese stammten von jeweils einem

Vertreter des Auftraggebers, der Servicestelle, der beiden Qualifizierungsträger, der Kursleiter beider Qualifizierungsträger sowie von zwei Vertretern der Gruppe der Eltern und drei Vertretern der Gruppe der Elternbegleiter. Ergänzt wurden diese Informationen durch die Stellungnahme eines Mitglieds des Evaluationsteams. Das Kompetenzteam gab keine Stellungnahme ab, weil es so eng mit dem Auftraggeber zusammenarbeitete, dass diese keine zusätzlichen Informationen erbracht hätte.

Die Auswahl der Stellungnehmer zielte nicht auf Repräsentativität ab, sondern sollte Personen einbeziehen, die Aussagen zur Evaluation treffen konnten. Daher wurden als Vertreter des Auftraggebers, der Servicestelle und der Träger gezielt Personen angesprochen, die mit der Evaluation bereits einige Berührungspunkte gehabt hatten und die Meta-Evaluatorin kannten. Auch ein Kursleiter[8] konnte auf diesem Weg erreicht werden, während der Kursleiter des anderen Trägers von dem Vertreter dieses Trägers vermittelt wurde. Aus den beiden Gruppen der Eltern und Elternbegleiter wurden zwei bzw. drei Vertreter einbezogen, weil einerseits diese Stakeholdergruppen deutlich größer waren als die übrigen und andererseits bei diesen Personen nicht mit vielen Schnittstellen zur Evaluation zu rechnen war, sodass sie vermutlich nur wenige Erfahrungen berichten und Einschätzungen abgeben konnten. Die Auswahl der Eltern erfolgte über die Liste der im Rahmen der Evaluation befragten Eltern. Vorausgewählt wurden Eltern, die sowohl am Prä- als auch am Posttest teilgenommen hatten, weil ihnen damit mehr Aussagen zur Evaluation möglich sein sollten. Unter diesen Probanden erfolgte eine Zufallsauswahl, wobei darauf geachtet wurde, dass der Posttest möglichst spät stattgefunden hatte, damit noch möglichst genaue Erinnerungen an die Befragungen abgerufen werden konnten. Seitens der Elternbegleiter wurden aus den Listen der im Rahmen der Evaluation befragten Personen diejenigen ausgewählt, die dem Evaluationsteam Eltern zur Befragung vermittelt hatten, um auch zu diesem Vermittlungsprozess Aussagen erhalten zu können. Zudem sollte eine Person vertreten sein, die den zweiten Bundeskongress Elternbegleitung besucht und somit auf jeden Fall Evaluationsergebnisse erfahren sowie die Kurzbroschüre erhalten hatte und dazu Aussagen treffen konnte. Auf Grundlage der Teilnehmerliste des

[8] Aus Anonymitätsgründen wird auf die Stellungnehmer stets in der männlichen Form (z. B. „Kursleiter“) Bezug genommen, unabhängig von ihrem tatsächlichen Geschlecht. Alternativ wird eine neutrale Form verwendet (z. B. Mitglied).

Bundeskongresses wurden somit im nächsten Schritt Teilnehmer und Nicht-Teilnehmer getrennt und aus beiden Gruppen per Zufallsauswahl Vertreter gezogen. Da sich allerdings zeigte, dass der Elternbegleiter, der auf der Teilnehmerliste gestanden hatte, doch nicht beim Kongress anwesend gewesen war, musste ein dritter Vertreter mit aufgenommen werden, um die Gruppe der Kongressteilnehmer abzudecken. Das stellungnehmende Mitglied des Evaluationsteams schließlich wurde danach ausgesucht, dass es über mehr Berufserfahrung verfügte als die Meta-Evaluatorin und außerdem am DJI arbeitete und somit einen anderen Blickwinkel auf den Evaluationsprozess mit einbringen konnte.

Wie bereits erwähnt, konnten die Stellungnehmer ihre Antworten schriftlich oder mündlich abgeben, wobei letzteres anhand von telefonischen Interviews mit Tonbandaufzeichnung durch die Meta-Evaluatorin umgesetzt wurde. Von den 12 Stellungnahmen wurden insgesamt drei schriftlich vorgenommen und neun mittels Telefoninterviews abgegeben. Aus Anonymitätsgründen wird nicht angegeben, welche Stakeholdervertreter schriftlich und welche mündlich Stellung nahmen, um etwaige Rückschlüsse aufgrund der Zitate in Abschnitt 4.2.2 zu verhindern. Die telefonischen Interviews wurden zunächst wörtlich transkribiert, um alle Stellungnahmen in Schriftform vorliegen zu haben. Die Analyse erfolgte nach pragmatischen Gesichtspunkten, indem die Meta-Evaluatorin die Inhalte der Stellungnahmen in Kategorien einordnete, welche sie deduktiv aus den Evaluationsstandards abgeleitet hatte. Die Auswertung orientierte sich somit an der Qualitativen Inhaltsanalyse (Mayring, 2010), jedoch wurde keine Kodierung durch zwei unabhängige Rater und somit auch keine Berechnung einer Beurteilerübereinstimmung zum Zweck der Reliabilitätsprüfung vorgenommen. Auf diese Auswertungsschritte wurde verzichtet, da sie einerseits den zeitlichen Rahmen der vorliegenden Arbeit überschritten hätten und auch keine Ressourcen dafür vorhanden waren, und sie andererseits auch nicht als unbedingt notwendig erachtet wurden. Denn es handelte sich nicht um qualitatives Material, das von Grund auf hätte erschlossen und interpretiert werden müssen, sondern um Antworten auf gezielte Fragen, sodass eine valide Interpretation auch durch eine Person möglich war.

Hinsichtlich der Stellungnahmen ist ergänzend zu erwähnen, dass sich zum Zeitpunkt ihrer Erhebung der Abschlussbericht noch in Bearbeitung befand und daher keine Fragen diesen Bericht betreffend

gestellt werden konnten. Evaluationsergebnisse waren einige Monate zuvor in Form der Kurzbroschüre veröffentlicht worden, sodass stattdessen nach dieser Publikation gefragt wurde. Den Stellungnehmern, denen diese Broschüre noch nicht bekannt war (vgl. Abschnitt 4.2.2.1.5), wurde sie nach der Stellungnahme zugeschickt (sofern hierfür Interesse bestand, was bei allen Betroffenen der Fall war).

Nach der Erhebung der Stellungnahmen erfolgte von Juni bis August 2015 ein zweiter Bewertungsdurchgang, bei welchem die Meta-Evaluatorin die Daten der Stellungnahmen in die Informationen des ersten Bewertungsdurchgangs integrierte und zudem die Bewertung ausführlich und möglichst nachvollziehbar schriftlich niederlegte. Diese verschriftlichte Bewertung ging sodann (im August 2015) an zwei der Stellungnehmer (Auftraggeber und Mitglied des Evaluationsteams), welche sie lasen und der Meta-Evaluatorin im Oktober 2015 eine Rückmeldung gaben (diese wird im Folgenden auch als 2. Stellungnahme des jeweiligen Stakeholders bezeichnet). Mit diesem Schritt verfolgte die Meta-Evaluatorin mehrere Ziele. So wollte sie der Subjektivität des eigenen Urteils entgegenwirken, indem sie ihre Bewertungen von zwei Stakeholdern prüfen und kommentieren ließ. Die Wahl fiel dabei auf das Mitglied des Evaluationsteams und den Auftraggeber, weil diese von allen Stakeholdern die meiste Einsicht in den gesamten Evaluationsprozess hatten. Das Gegenlesen durch den Auftraggeber hatte zudem die Funktion der Absicherung, damit in den Bewertungen keine Angaben enthalten waren, welche nicht zur Veröffentlichung bestimmt waren. Und schließlich hatten der Meta-Evaluatorin noch vereinzelte Informationen gefehlt, die sie auf diesem Wege nachfragen konnte. Es ist anzumerken, dass, um die Subjektivität noch mehr zu reduzieren, Rückmeldungen weiterer Stakeholder zu den verschriftlichten Bewertungen sinnvoll gewesen wären, da sie weitere Perspektiven eingebracht hätten. Dies konnte allerdings aus zeitlichen Gründen nicht realisiert werden, weil bereits die Rückmeldungen von Auftraggeber und Mitglied des Evaluationsteams mehrere Wochen gedauert hatten, da die Evaluationsbewertung etwa 70 Seiten umfasste (vgl. Abschnitt 4.2.2).

Die Kommentare des Auftraggebers und des Mitglieds des Evaluationsteams wurden in der Folge in die Evaluationsbewertung eingearbeitet. Dabei wurde so vorgegangen, dass sämtliche Anregungen zunächst gründlich geprüft und dabei auch beide Stellungnahmen

miteinander abgeglichen wurden. Inhaltliche Ergänzungen und Formulierungsvorschläge konnten weitgehend übernommen werden, zumal einige bislang fehlende Daten von der Meta-Evaluatorin direkt erfragt worden waren. Bei den Änderungsvorschlägen, welche die Bewertungen selbst betrafen, übernahm die Meta-Evaluatorin einige Änderungen, die von beiden Stakeholdern angemerkt worden waren. Einige Anregungen, die von nur einem Stakeholder vorgebracht worden waren, wurden nach Abwägung mit dem eigenen Urteil entweder ebenfalls in die Bewertung integriert oder, sofern die Meta-Evaluatorin sich dieser Meinung nicht anschließen konnte, als Statement des betreffenden Stakeholders gekennzeichnet und angefügt.

Einen Aspekt, der von beiden Stellungnehmern angemerkt wurde und auf den hier hingewiesen werden soll, stellte die zu geringe Einordnung der vorliegenden Evaluation als Auftragsforschung dar. Die Meta-Evaluatorin hatte sich bei ihrer Bewertung eng an den Evaluationsstandards orientiert, welche viel Einbezug der Beteiligten empfehlen. Dieses Vorgehen führte zu einigen Abzügen insbesondere bei den Nützlichkeitsstandards, was den Stellungnehmern nicht nachvollziehbar erschien in Anbetracht der Tatsache, dass bei dieser Evaluation ein auftraggeberorientierter Ansatz verfolgt wurde und sie keine partizipative Ausrichtung hatte. Die Meta-Evaluatorin entschied sich, dieser Argumentation folgend, bei den betreffenden Standards auf die Auftraggeberorientierung der Evaluation hinzuweisen und ihre Bewertung entsprechend einzuordnen. Die Frage, inwiefern die Evaluationsstandards für verschiedene Evaluationsansätze passend sind, wird in der Diskussion erörtert (Abschnitt 5.2.1).

Hinsichtlich der verwendeten Methode ist zu bemerken, dass diese viele Ähnlichkeiten zum Vorgehen von Widmer (1996) aufweist. So wurden Unterlagen ausgewertet und ergänzend Stellungnahmen von Beteiligten zu einem Fragenkatalog erhoben. Die verschriftlichte Bewertung wurde zwei Stakeholdern mit der Bitte um Rückmeldung vorgelegt und diese Rückmeldung anschließend eingearbeitet. Zudem wurde aus Gründen der Nachvollziehbarkeit eine ausführliche Darlegung der Bewertungen gewählt. Interessanterweise waren die ausführliche, nachvollziehbare Schilderung sowie die Auswertung der vorhandenen Unterlagen die einzigen Schritte, die vorab analog zu Widmer geplant gewesen waren, bei denen also eine explizite Orientierung an dessen Methode erfolgte. Das weitere Vorgehen (die Erhebung

der ersten und auch der zweiten Stellungnahmen) erwies sich erst während der Evaluationsbewertung als erforderlich und nützlich. Somit ergab sich sozusagen induktiv im Bewertungsprozess eine Methode, die schließlich der von Widmer sehr ähnelte.

Nach dieser detaillierten Schilderung des Meta-Evaluationsprozesses wird im folgenden Abschnitt das Ergebnis, also die Evaluationsbewertung, dargestellt.

4.2.2 Bewertung der Evaluation des Bundesprogramms „Elternchance ist Kinderchance" anhand von Evaluationsstandards

Die Bewertung der Evaluation des Bundesprogramms „Elternchance ist Kinderchance" folgt der Gliederung der DeGEval-Standards (vgl. Abschnitt 2.2.3). Dabei werden der Reihe nach die Einzelstandards der vier Standardgruppen *Nützlichkeit*, *Durchführbarkeit*, *Fairness* und *Genauigkeit* behandelt[9]. Die Bewertung jedes Einzelstandards ist nach folgendem Schema aufgebaut: Zunächst werden die Anforderungen genannt, die die DeGEval-Richtlinien (DeGEval, 2008b) zum jeweiligen Standard enthalten. Diese werden um Erläuterungen des JCSEE-Handbuchs erweitert (JCSEE, 2006). Nachfolgend werden Informationen aus dem Evaluationsprojekt berichtet, die Aufschluss über die (Nicht-)Einhaltung des Standards geben. Sie werden um Daten aus den Stellungnahmen ergänzt, sofern diese Anhaltspunkte zum jeweiligen Standard enthielten. Die zusammengetragenen Informationen werden schließlich resümiert und zu einer Bewertung des Standards integriert.

Zum besseren Verständnis der folgenden Abschnitte seien einige Hinweise gegeben. Die Erläuterungen der DeGEval-Standards werden für jeden Standard vollständig abgebildet. Die entsprechenden Hinweise des JCSEE-Handbuchs ebenso vollständig darzustellen, würde jedoch den Rahmen dieser Arbeit übersteigen. Es werden daher für jeden Standard nur diejenigen Handbuch-Hinweise geschildert, die sich als für die vorliegende Evaluation passend herausgestellt hatten (vgl. Abschnitt 4.2.1). Auf die im Handbuch enthaltenen Fallbeispiele

[9] Im Interesse der besseren Lesbarkeit werden in den Abschnitten 4.2.2.1 bis 4.2.2.4 die Kurzbeschreibungen der Einzelstandards sowie der vier Standardgruppen nochmals angeführt. Diese sind, ebenso wie in Abschnitt 2.2.3, wörtlich (Schreibfehler ausgenommen) den Standards für Evaluation der DeGEval (2008b, S. 23-38) entnommen.

wird nicht eingegangen, da sie keine Relevanz für diesen Abschnitt haben.

Die zum Evaluationsprojekt berichteten Daten basieren auf den im vorigen Abschnitt dargestellten Quellen. Einige dieser Informationen wurden bereits im Rahmen der Projektbeschreibung im Abschnitt 4.1 vermittelt und werden hier ggf. um weitere Angaben ergänzt.

Hinsichtlich der Stellungnahmen ist zu erwähnen, dass, wenn im Folgenden nur von „Stellungnahme“ die Rede ist, hiermit die Antworten der Stakeholder zum Fragenkatalog gemeint sind. Die Rückmeldungen von Auftraggeber und Mitglied des Evaluationsteams zur schriftlichen Bewertung werden in Abgrenzung dazu als „2. Stellungnahme“ dieser beiden Beteiligten bezeichnet. Die Reihenfolge, in welcher die Stellungnahmen verschiedener Stakeholder angegeben werden, ist nicht als Hinweis auf eine mögliche Priorisierung der Stellungnahmen zu verstehen. Die Angaben, die im Rahmen der Stellungnahmen (auch der zweiten Stellungnahmen) gemacht wurden, werden im Konjunktiv berichtet, um deutlich zu machen, dass es sich um Aussagen der Beteiligten handelt. Eine Ausnahme stellen Informationen inhaltlicher Art dar, die keine subjektive Einschätzung der betreffenden Person widerspiegeln. Solche Informationen, z. B. Daten oder Vorgänge das Bundesprogramm betreffend, werden im Rahmen der Projektdaten ohne Nennung der Quelle berichtet, da sie keine Meinung des Stellungnehmers enthalten. Dies betrifft insbesondere die beiden Stellungnahmen des Mitglieds des Evaluationsteams, in welchen viele ergänzende inhaltliche Informationen gegeben wurden.

Abschließend sei darauf hingewiesen, dass in den folgenden Abschnitten lediglich die Bewertung der Evaluation entlang der Evaluationsstandards erfolgt. Die dabei angeschnittenen Herausforderungen werden ebenso wie die angewendete Methodik in Kapitel 5 diskutiert, wo auch Empfehlungen abgeleitet werden.

4.2.2.1 Nützlichkeitsstandards

Die Nützlichkeitsstandards sollen sicherstellen, dass die Evaluation sich an den geklärten Evaluationszwecken sowie am Informationsbedarf der vorgesehenen Nutzer und Nutzerinnen ausrichtet.

4.2.2.1.1 Identifizierung der Beteiligten und Betroffenen (N1)

> Die am Evaluationsgegenstand beteiligten oder von ihm betroffenen Personen bzw. Personengruppen sollen identifiziert werden, damit deren Interessen geklärt und so weit wie möglich bei der Anlage der Evaluation berücksichtigt werden können.

Diese DeGEval-Richtlinie umfasst zwei Aspekte: Die Identifizierung der Beteiligten und Betroffenen einerseits sowie die Ermittlung und Berücksichtigung ihrer Interessen bei der Evaluation andererseits. Die DeGEval-Standards geben Hilfestellung für die Identifizierung der Beteiligten und Betroffenen, indem Hinweise zu deren Ermittlung angeführt werden und auch auf die Beachtung möglicher durch das Evaluandum benachteiligter oder davon ausgeschlossener Personen verwiesen wird. Weiterhin wird ausgeführt, dass die identifizierten Stakeholder den Evaluationszwecken entsprechend bei der Planung und Durchführung der Evaluation beteiligt werden sollen, sofern dies finanziell und zeitlich möglich ist. Zudem wird betont, dass die Informationsbedürfnisse der Beteiligten geklärt und berücksichtigt werden sollen. Die Ausrichtung einer Evaluation an den Stakeholdern und ihren Informationsbedürfnissen wird als sehr bedeutend bewertet, was die spätere Nutzung der Evaluationsergebnisse betrifft. Hierzu wird empfohlen, dass auch weitere Gruppen beachtet werden sollen, für welche die Evaluationsergebnisse relevant sein könnten (z. B. Fachöffentlichkeit).

Das JCSEE-Handbuch enthält zusätzlich veranschaulichende Beispiele für denkbare Beteiligte. Es wird zudem der Hinweis gegeben, dass die angemessene Beteiligung von weniger einflussreichen Gruppen (z. B. sprachliche Minderheiten) eine besondere Herausforderung darstellen kann. Des Weiteren werden konkretere Hinweise zur Identifizierung der Beteiligten und ihrer Informationsbedürfnisse gegeben sowie der Aufbau realistischer Erwartungen bei den Stakeholdern angesprochen. Auch wird empfohlen, die Bedeutung der Beteiligten mit dem Auftraggeber gemeinsam zu klären und darauf aufbauend die Datenerhebung und Berichtlegung auszurichten.

Zunächst sei in Erinnerung gerufen, dass es sich bei dieser Evaluation um eine auftraggeberorientierte handelte, dass also der Auftraggeber den zentralen Stakeholder darstellte, an dessen Weisungen das

Evaluationsteam vertraglich gebunden war. Zudem war die Evaluation in erster Linie darauf ausgelegt, die Bedarfe des Auftraggebers zu erfüllen, was die Beteiligung anderer Stakeholder beeinflusste.

Bei der vorliegenden Evaluation wurden verschiedene Beteiligte und Betroffene identifiziert (vgl. Abschnitt 4.1.2) und in unterschiedlichem Maße in den Evaluationsprozess miteinbezogen. Das BMFSFJ als Auftraggeber stellte die Institution dar, mit welcher das Evaluationsteam am längsten und intensivsten in Kontakt stand. Es gab zudem das sogenannte „Kompetenzteam Wissenschaft", welches aus Wissenschaftlern bestand, die als Beratungsinstanz mit dem BMFSFJ zusammenarbeiteten. Sowohl das BMFSFJ als auch das Kompetenzteam waren in alle Schritte der Evaluation (insbesondere Planung der Datenerhebung und Berichtlegung) eng eingebunden worden, was der Anlage der Evaluation als auftraggeberorientiert entsprach.

Am Bundesprogramm beteiligt war weiterhin die Servicestelle Elternchance, die für die fachlich-inhaltliche und finanziell-administrative Begleitung und Umsetzung des Programms im Auftrag des BMFSFJ zuständig war. Im Rahmen des Monitorings erhob die Servicestelle zudem selbst Daten bei Qualifizierungsteilnehmern sowie bei den Einrichtungen *Elternbegleitung Plus*. Hinsichtlich der Planung der Datenerhebung hatten sich Servicestelle und Evaluatoren abgestimmt, damit sich in ihren Befragungen keine Überschneidungen ergaben und somit Daten nicht doppelt erfragt wurden. Da die von der Servicestelle erhobenen Daten auch vom Evaluationsteam genutzt wurden, fand auch diesbezüglich ein Austausch statt. Zudem wurde die Servicestelle über (Zwischen-)Ergebnisse im Rahmen von Workshops informiert.

Einen weiteren Stakeholder stellte die Impulsgruppe dar, ein wissenschaftlicher Beirat aus Experten der Evaluationsforschung sowie der Erwachsenenbildung und Familienforschung, der die Evaluation begleitete. Mit diesem Beirat fand eine Zusammenarbeit in Form von Austausch und Diskussion zu verschiedenen Zeitpunkten im Evaluationsverlauf statt, bei denen die Impulsgruppe, ihrem Namen getreu, wichtige Impulse setzte.

Die beiden Qualifizierungsträger, welche das Bundesprogramm durchführten, waren weitere Stakeholder. Diese waren an der Planung der Erhebungsinstrumente nicht beteiligt gewesen (eine Ausnahme stellten die zentralen Inhalte der Qualifizierung dar, zu welchen es eine Abstimmung gegeben hatte), erhielten jedoch Informationen zu

diesen. Außerdem wurden ihre Perspektiven so weit wie möglich berücksichtigt, indem bei der Anlage der Erhebungsinstrumente versucht wurde, den beiden unterschiedlichen Ausbildungscurricula gerecht zu werden. Im Rahmen der Datenerhebung fand eine enge Zusammenarbeit zwischen dem DJI und den Trägern statt, die insbesondere den Zugang zu den Probanden umfasste. Positiv hervorzuheben ist dabei das Engagement der Träger, welche die Evaluation sehr unterstützten, indem sie beispielsweise in den Qualifizierungskursen für die Teilnahme an der Befragung warben. Den Trägern wurden weiterhin im Rahmen von Workshops aktuelle Zwischenergebnisse präsentiert. Die sich daran anschließenden Diskussionen ermöglichten einen Einbezug der Sichtweisen dieser Stakeholder in die Interpretation der Daten. Die Dozenten, die als Kursleiter die Qualifizierungen zu Elternbegleitern durchführten, waren ebenfalls Beteiligte. Diese waren nicht direkt in die einzelnen Schritte der Evaluation einbezogen worden, da sie zu der Instanz der Qualifizierungsträger gehörten und davon auszugehen war, dass ihnen wichtige Informationen von dieser Stelle weitergegeben wurden. Auf der Ebene der Kursleiter wurden zudem in Form von Experteninterviews Daten erhoben, sodass ihre Perspektive in diesem Rahmen Berücksichtigung fand.

Die von den Dozenten zu Elternbegleitern qualifizierten Personen stellten ebenso wie die Fachkräfte der Kontrollgruppe sowie die von beiden Gruppen begleiteten Eltern weitere Beteiligte dar. Beide Gruppen wurden im Rahmen der Datenerhebung befragt. Bereits bei der Konstruktion der Erhebungsinstrumente fand im Rahmen einer Pilotstudie ein Einbezug dieser Beteiligten statt, deren Rückmeldungen zur Anpassung der Fragebögen verwendet wurden. Es wurden sowohl quantitative als auch qualitative Daten erhoben, um vielfältige Informationen zu erhalten und die Situation der beiden Gruppen angemessen berücksichtigen und im Kontext interpretieren zu können. Zudem wurde die Datenerhebung insbesondere auf Elternebene auf die Zielgruppen des Bundesprogramms (bildungsferne und armutsgefährdete Personen sowie Menschen mit Migrationshintergrund) ausgerichtet, indem leicht verständliche Formulierungen gewählt, Wertgutscheine als Anreize verwendet und die Interviews auch in Fremdsprachen angeboten wurden. Es wurden also zusätzliche Bemühungen unternommen, um die Zielgruppen des Programms erreichen und angemessen beteiligen zu können.

Anhand der bis hierhin beschriebenen Informationen lässt sich die Schlussfolgerung ziehen, dass die Identifizierung der Beteiligten und Betroffenen gut gelungen war. Es ist nicht absehbar, dass wichtige Gruppen übersehen wurden und in die Erreichung bestimmter Gruppen wurden sogar besondere Anstrengungen investiert. Auch die Beteiligung der Stakeholder sowie die Berücksichtigung ihrer Interessen erscheinen als gut gelungen, insbesondere im Rahmen der gegebenen finanziellen und zeitlichen Ressourcen sowie in Anbetracht der Tatsache, dass es sich um eine auftraggeberorientierte Evaluation handelte. Um zu prüfen, ob diese Einschätzung mit der Sichtweise der Beteiligten übereinstimmt und diese sich angemessen in die Evaluation einbezogen fühlten, wurden die Stellungnahmen herangezogen, in welchen dieser Aspekt erörtert wurde.

Generell können die Äußerungen der Befragten dahingehend zusammengefasst werden, dass alle ihre Beteiligung am Evaluationsprozess als gut bzw. sehr gut einschätzten. Jedoch wurden auch einige Bereiche genannt, in denen verschiedene Stakeholder sich mehr Beteiligung gewünscht hätten. Am häufigsten bemängelt wurde ein Defizit in der Ergebnisverbreitung. So waren die Eltern, die Elternbegleiter sowie teilweise die Kursleiter zum Zeitpunkt ihrer Stellungnahmen kaum bzw. gar nicht über (Zwischen-)Ergebnisse der Evaluation informiert worden. Dieser Aspekt wird in der vorliegenden Arbeit im Rahmen der Rechtzeitigkeit des Berichts beim Standard N7 bzw. im Rahmen der Ergebnisnutzung beim Standard N8 erörtert und daher an dieser Stelle nicht näher ausgeführt.

Im Hinblick auf die Planung der Erhebungsinstrumente sowie die Datenerhebung ergab sich, dass mehrere Stakeholder in diese Phase gerne mehr einbezogen worden wären. Dies betraf zum einen die Gruppe der Qualifizierungsträger, deren einer Vertreter sowie der Kursleiter desselben Trägers deutlich machten, dass die Erhebungsinstrumente zu wenig Rückmeldung über die Stärken und Schwächen der Qualifizierung ermöglicht hätten, welche jedoch von besonderem Interesse gewesen wären. Somit habe die Evaluation die für diesen Stakeholder wichtigen Fragen nicht angemessen beantworten können und es seien daher eigene Erhebungen notwendig gewesen. Mit einer besseren Beteiligung bei der Planung der Erhebungsinstrumente hätte man dies womöglich umgehen können. Der Vertreter dieses Trägers bemerkte außerdem, dass es vielleicht sinnvoll gewesen wäre, den Dozenten frühzeitig detaillierte Informationen zukommen zu lassen,

um die Bereitschaft der Qualifizierungsteilnehmer zur Mitarbeit zu erhöhen. Zum anderen erwähnte der Vertreter der Servicestelle, dass mehr Absprache bezüglich der Durchführung der Erhebungen sinnvoll gewesen wäre. Die Servicestelle erhob in den Einrichtungen *Elternbegleitung Plus* eigene Daten und es kam im Verlauf der Evaluation vor, dass diese Erhebungen zeitlich sehr nah an den Befragungen des Evaluationsteams lagen, was mit entsprechenden Belastungen für die Befragten einherging. Daher wurde im Rahmen der Stellungnahme einerseits der Wunsch nach mehr Abstimmung geäußert und andererseits angeregt, dass zu Beginn der Evaluation ein Informationstreffen bezüglich Planung und Umsetzung mit allen Stakeholdern sinnvoll gewesen wäre. Zu erwähnen ist hierbei, dass es durchaus eine Absprache zwischen Evaluatoren und Servicestelle gegeben hatte, die auch das BMFSFJ den beiden Parteien empfohlen hatte. Allerdings konnten dennoch die Erhebungen nur bedingt verschoben bzw. aufeinander abgestimmt werden, da jeweils Zeitpläne eingehalten werden mussten. Zudem nahmen die Einrichtungen *Elternbegleitung Plus* nicht nur an diesen beiden Erhebungen teil, sondern wurden zusätzlich durch zwei weitere Befragungen belastet, die von Servicestelle und Kompetenzteam für das BMFSFJ durchgeführt wurden. Dies war einer der Gründe, weshalb die im Rahmen der Evaluation ursprünglich geplante dritte Erhebung des Moduls 2 nicht realisiert wurde.

Die Idee einer direkten Beteiligung an der Entwicklung der Erhebungsinstrumente wurde überwiegend abgelehnt, da die Befragten die Kompetenz für diesen Schritt beim Evaluationsteam und nicht bei sich selbst oder ihrer Stakeholdergruppe sahen. Den Vorschlag, Elternbegleiter bei der Konstruktion des Elternfragebogens zu Rate zu ziehen, befürworteten hingegen zwei der drei befragten Elternbegleiter sowie ein Kursleiter. Dies böte die Möglichkeit, die inhaltlichen und methodischen Kompetenzen der Evaluatoren mit den Praxiseinsichten der Elternbegleiter zu verknüpfen.

Ein weiterer Punkt, der von einem Vertreter der Qualifizierungsträger angemerkt wurde, war die unterschiedlich ausgeprägte Kooperation mit den beiden Teilen des Evaluationsteams. So sei die Zusammenarbeit mit dem DJI sehr eng gewesen, mit der FAU hingegen weniger, was jedoch durch die unterschiedlichen Aufgabenbereiche von DJI und FAU erklärt werden könne und auch nachvollziehbar sei. Dennoch wäre es im Nachhinein betrachtet wünschenswert gewesen, auch mit der FAU etwas mehr Austausch zu haben. Allerdings sei dies kein

Kritikpunkt, zumal sich das Projekt im Verlauf der Evaluation entwickelt habe und nicht von Beginn an klar gewesen sei, wer welche Rolle spielen werde. Daher sei es in Ordnung, wie es bei dieser Evaluation gehandhabt worden sei, jedoch könne man diesen Aspekt bei einer künftigen Evaluation verbessern. Zudem habe die Evaluation nun positive Ergebnisse erbracht und die Qualifizierungsträger in ihrer Arbeit weitgehend bestätigt. Wenn dies jedoch anders gewesen wäre, dann hätte man vielleicht darauf bestanden, zu wenig einbezogen worden zu sein. Jedoch bestehe nun einmal dieses Spannungsfeld zwischen wissenschaftlicher Herangehensweise und zu großer Nähe zu den Befragten, wofür man auch Verständnis habe.

Mit diesem Aspekt in Verbindung stehend äußerte die gleiche Person den Wunsch nach mehr Kommunikation auch dahingehend, dass man in die Entstehung der Kurzbroschüre nicht einbezogen worden sei. Diese sei auf einmal fertig gewesen, ohne dass es zuvor eine Beteiligung oder Information gegeben habe. Durch mehr Austausch könne auch der Praxisanschluss deutlich verbessert werden.

Der Auftraggeber betonte in seiner zweiten Stellungnahme, dass die Wünsche einiger Stakeholder nach mehr Beteiligung am Evaluationsprozess in Bezug zu der Tatsache zu setzen seien, dass es sich bei der vorliegenden Studie um Auftragsforschung gehandelt habe und damit nicht um eine Kooperation von Forschung und Praxis.

Werden die dargestellten Informationen zusammengenommen betrachtet, scheint insgesamt sowohl die Identifizierung der Stakeholder als auch ihre Beteiligung am Evaluationsprozess weitgehend gelungen zu sein. Bezüglich der Beteiligung ergaben die Stellungnahmen kleinere Schwächen der vorliegenden Evaluation aus Sicht einiger Beteiligter, die in der Diskussion aufgegriffen werden. Die genannten Wünsche einiger Stakeholder nach mehr Beteiligung in verschiedenen Bereichen müssen allerdings gegen die zeitlichen und finanziellen Ressourcen der Evaluierenden abgewogen werden. Zudem ist zu beachten, dass es sich beim vorliegenden Projekt um eine auftraggeberorientierte Evaluation handelte und somit eine möglichst hohe Beteiligung aller Stakeholder nicht (wie beim partizipativen Ansatz, vgl. Abschnitt 2.1.3.5) im Mittelpunkt stand. Daher ergibt sich für den Standard N1 folgende Bewertung:

⇨ Standard N1 wurde überwiegend erfüllt.

4.2.2.1.2 Klärung der Evaluationszwecke (N2)

> Es soll deutlich bestimmt sein, welche Zwecke mit der Evaluation verfolgt werden, sodass die Beteiligten und Betroffenen Position dazu beziehen können und das Evaluationsteam einen klaren Arbeitsauftrag verfolgen kann.

Den DeGEval-Standards zufolge sollen die Evaluationszwecke frühzeitig angesprochen und im Evaluationsverlauf verhandelt und schriftlich festgelegt werden. Sie sollen zudem soweit möglich den Beteiligten und Betroffenen zugänglich gemacht werden, damit diese an der Planung der Evaluation teilhaben können. Die Klärung der Evaluationszwecke stellt eine wichtige Grundlage für die Ergebnisverbreitung und -nutzung dar. Die DeGEval-Standards enthalten weiterhin eine Übersicht über drei idealtypische Evaluationszwecke (Optimierung, Entscheidung und Erkenntnisgewinn). Es wird davor gewarnt, mehrere Hauptzwecke gleichrangig und zur gleichen Zeit zu verfolgen, da dies die Nutzung beeinträchtigen kann. Für den Fall, dass mehrere Hauptzwecke vorliegen, wird empfohlen, diese zu priorisieren und nacheinander oder durch verschiedene Evaluationsteams zu bearbeiten.

Da dieser Standard in den JCSEE-Richtlinien nicht enthalten ist, kann ein Abgleich mit dem Handbuch nicht stattfinden.

Bei der Evaluation von „Elternchance ist Kinderchance“ war der Zweck der Evaluation bereits in der Auftragsbeschreibung im Rahmen der Ausschreibung enthalten und dadurch schriftlich erfasst. Er bestand in der Optimierung des Bundesprogramms, welches als offener und diskursiver Prozess bezeichnet wurde und zu Optimierungszwecken formativ und summativ evaluiert werden sollte. Diese Zwecksetzung war somit klar beschrieben und schriftlich festgelegt, sodass dieser Teil der Standardanforderungen als erfüllt bewertet werden kann.

Jedoch wurde in der Ausschreibung ein weiterer Punkt angesprochen, welcher sich im Verlauf der Evaluation herausfordernd gestaltete. Danach sollte in der Evaluation die Tatsache Berücksichtigung finden, dass die Qualifizierungen zu Elternbegleitern anhand zweier verschiedener Ausbildungscurricula stattfanden. Im Zusammenhang mit der Konstruktion der Erhebungsinstrumente wurde indessen zwischen Auftraggeber und Evaluationsteam mündlich vereinbart, dass Unterschiede zwischen den Curricula nicht untersucht werden sollten.

Im Verlauf der Evaluation zeigte sich jedoch, dass ein Vergleich der beiden Curricula sinnvoll wäre, um deren Vor- und Nachteile untersuchen und dadurch wichtige Informationen für die Fortsetzung des Bundesprogramms erhalten zu können. Daher äußerte der Auftraggeber, auf Empfehlung des DJI hin, schließlich doch den Wunsch nach vergleichenden Ergebnissen zu den beiden Curricula. Jene Ergebnisse waren nur für interne Zwecke gedacht und sollten keinen Eingang in den Abschlussbericht der Evaluation finden. Als schwierig erwies sich hierbei die Kommunikation mit den beiden Qualifizierungsträgern. Diese hatten aufgrund der Ausschreibung zunächst die Erwartung, dass die Evaluation ihnen eine Rückmeldung zu ihrem spezifischen Curriculum liefern würde. Nach der genannten mündlichen Vereinbarung wurde ihnen gegenüber kommuniziert, dass Unterschiede zwischen den Curricula nicht untersucht würden, was schließlich aber doch getan wurde.

Zur Beantwortung der Frage, ob mehrere Zwecke verfolgt wurden, ist die erste Stellungnahme des Auftraggebers hilfreich. Hier wurde neben der Optimierung auch benannt, dass die Evaluation für die Weiterführung und -entwicklung des Bundesprogramms werben sollte. Dies weist darauf hin, dass mit positiven Ergebnissen der Evaluation gerechnet wurde. Allerdings ist dies nicht als Hauptzweck im Sinne des Standards N2 zu bewerten, der Ressourcen des Evaluationsteams beansprucht hätte, sondern eher als strategischer Zweck, der nebenher verfolgt wurde. Dieser wurde insbesondere bei der Berichtlegung deutlich, als von Seiten des Auftraggebers sehr auf das dort verwendete Wording geachtet wurde (vgl. Standard D2). Wie der Auftraggeber in seiner zweiten Stellungnahme anmerkte, sei auch hier die Einordnung der vorliegenden Evaluation als Auftragsforschung zu beachten, da Auftragsforschung neben der Zielsetzung der Qualitätskontrolle auch strategische Ziele verfolge.

Insgesamt war zumindest der offizielle Evaluationszweck geklärt und schriftlich fixiert und es wurden nicht mehrere Hauptzwecke parallel verfolgt. Da jedoch bei der Kommunikation der Evaluationszwecke an die Beteiligten aus dem oben erwähnten Grund Abstriche zu verzeichnen sind, kann dieser Evaluationsstandard nur als überwiegend erfüllt gelten.

⇨ Standard N2 wurde überwiegend erfüllt.

4.2.2.1.3 Glaubwürdigkeit und Kompetenz des Evaluators/der Evaluatorin (N3)

> Wer Evaluationen durchführt, soll persönlich glaubwürdig sowie methodisch und fachlich kompetent sein, damit bei den Evaluationsergebnissen ein Höchstmaß an Glaubwürdigkeit und Akzeptanz erreicht wird.

Wie die DeGEval-Standards herausstellen, ist die Glaubwürdigkeit der Evaluierenden ein wichtiger Faktor bei der Durchführung und Wirksamkeit einer Evaluation. Um sicherzustellen, dass sie von den Stakeholdern als glaubwürdig wahrgenommen werden, sind bei den Evaluierenden fachliche und methodische Kompetenzen, Integrität und Unabhängigkeit sowie soziale und kommunikative Kompetenzen unabdingbar. Diese können auch durch die Zusammenstellung eines Evaluationsteams erreicht werden.

Während die deutschen Richtlinien nur beschreiben, welche Kompetenzen zu einem Evaluationsteam gehören, welches das Kriterium der Glaubwürdigkeit erfüllt, gibt das Handbuch des JCSEE zusätzlich viele Hilfestellungen, auf welche Weise die Glaubwürdigkeit erhöht werden kann und welche Handlungen und Maßnahmen sie hingegen einschränken können. Diese können aus Platzgründen an dieser Stelle nicht sämtlich einzeln aufgeführt werden, stellen jedoch insgesamt eine wichtige Ergänzung zu den DeGEval-Richtlinien dar. So wird unter anderem darauf hingewiesen, dass die Evaluation von anderen Evaluatoren überprüft werden kann, die vom Auftraggeber als glaubwürdig eingestuft werden (Meta-Evaluation). Zudem wird auch davor gewarnt, zu viele Mittel in das Erreichen von Glaubwürdigkeit und Akzeptanz zu investieren.

Wie von den Standards empfohlen, wurden bei der Evaluation von „Elternchance ist Kinderchance“ die erforderlichen Kompetenzen durch die Bildung eines Evaluationsteams sichergestellt. Dieses setzte sich aus den beiden Institutionen DJI und FAU zusammen, wobei mit dem DJI als anerkannter Forschungseinrichtung auf dem Gebiet der Kindheits-, Jugend- und Familienforschung und dem Lehrstuhl für Psychologische Diagnostik, Methodenlehre und Rechtspsychologie der FAU jeweils sowohl fachliche als auch methodische Kompetenzen gewährleistet waren. So dürften bereits die Kompetenzprofile der

beiden Projektleiter Prof. Dr. Sabine Walper (DJI) und Prof. Dr. Mark Stemmler (FAU) für eine hohe Glaubwürdigkeit gesorgt haben. Weiterhin bestand das Team aus Psychologen, Sozialwissenschaftlern und Pädagogen, die neben breitem fachlichem Wissen über fundierte methodische Kenntnisse im quantitativen und qualitativen Bereich verfügten. Insbesondere die multidisziplinäre Zusammensetzung des Evaluationsteams erwies sich als sehr gewinnbringend, da die unterschiedlichen Hintergründe und Perspektiven im Projektverlauf zu wertvollen und fruchtbaren Diskussionen beitrugen.

Als problematisch stellte sich im Projektverlauf die hohe Personalfluktuation im Evaluationsteam dar (vgl. Abschnitt 4.1.2). So verblieben von ursprünglich sechs Projektmitarbeitern lediglich zwei bis zum Ende der Evaluation im Team, jeweils eine Person bei DJI und FAU. Auf Seiten des DJI schieden insgesamt drei von vier Mitarbeiterinnen aus und wurden durch zwei neue Teammitglieder ersetzt. Auch bei der FAU verließen ein Mitarbeiter sowie die Projektkoordinatorin das Team, weshalb eine neue Mitarbeiterin hinzukam und die koordinativen Aufgaben umverteilt wurden. Die hohe Fluktuation resultierte in einem erhöhten Einarbeitungsaufwand für das Evaluationsteam, der sich jedoch nicht auf die Glaubwürdigkeit des Teams auswirkte, wie unten anhand der Stellungnahmen gezeigt wird.

Die Evaluation wurde vom Kompetenzteam Wissenschaft begleitet, welches in jeden Evaluationsschritt eingebunden war und hierzu Rückmeldungen gab. Zusammen mit der Begleitung durch die wissenschaftliche Impulsgruppe kam dies einer proaktiven, formativen Meta-Evaluation gleich, während die vorliegende Arbeit eine retroaktive, summative Meta-Evaluation darstellt. Die Tatsache, dass die Arbeit der Evaluatoren auf diese Weise bewertet wurde und wird, trägt ebenfalls zu ihrer Glaubwürdigkeit bei.

Das Evaluationsteam führte viele Aktivitäten durch, um sich bei den verschiedenen Beteiligten Glaubwürdigkeit zu verschaffen. So war es bereits beim Vergabeverfahren gelungen, den Auftraggeber von den Kompetenzen zu überzeugen und auch im weiteren Verlauf fand ein regelmäßiger Austausch über die Projektfortschritte statt. Dabei wurden seitens der Evaluatoren stets große Anstrengungen unternommen, um beispielsweise Projekttreffen bestmöglich vorzubereiten. Auch dem Austausch mit den Qualifizierungsträgern wurde eine hohe Bedeutung beigemessen, weshalb Workshops mit den Evaluatoren organisiert wurden und auch Mitglieder des Evaluationsteams an Veranstaltungen

der Träger teilnahmen. Den befragten Fachkräften und Eltern wurde die Ansprechbarkeit der Evaluatoren (z. B. im Falle von Fragen) durch die Weitergabe von deren Kontaktdaten signalisiert. Zudem wurde zur Verbesserung der Erhebungsinstrumente eine Pilotstudie durchgeführt, im Rahmen derer explizit die Rückmeldung der teilnehmenden Elternbegleiter und Eltern erbeten wurde. Das Evaluationsteam nahm zudem am 1. und 2. Bundeskongress Elternbegleitung des BMFSFJ teil, wo es für einen Austausch mit den Anwesenden, darunter viele Elternbegleiter, zur Verfügung stand. Auch ein Vortrag bei der Volkshochschule Erlangen ergab die Möglichkeit zur Diskussion mit anwesenden Elternbegleitern.

Eine noch höhere Glaubwürdigkeit hätte womöglich durch eine persönlichere Rekrutierungsstrategie der Stichproben der Fachkräfte und Eltern erreicht werden können. Die Elternbegleiter und die Kontrollgruppen-Fachkräfte wurden per E-Mail angeschrieben und die Eltern von ihren Fachkräften vermittelt und dann telefonisch von der FAU kontaktiert. Eine Vorstellung des Projekts und die Bitte um Teilnahme daran durch das Evaluationsteam persönlich vor Ort hätten vielleicht die Teilnahmequoten sowie die Glaubwürdigkeit verbessert. Aus logistischen und finanziellen sowie zeitlichen Gründen war dies jedoch nicht möglich.

Die vielfältigen Bemühungen des Evaluationsteams resultierten tatsächlich in einer hohen wahrgenommenen Glaubwürdigkeit und Kompetenz durch die Beteiligten, wie die Stellungnahmen belegen. So zeigten sich sämtliche Stakeholder zufrieden oder sehr zufrieden hinsichtlich der Zusammenarbeit mit den Evaluatoren und äußerten weder bezüglich der Kompetenz noch in Bezug auf die Glaubwürdigkeit Bedenken. Als einziger kritischer Aspekt, der jedoch nicht direkt mit der Glaubwürdigkeit oder Kompetenz in Verbindung stand, wurde vom Kursleiter eines Trägers die etwas unübersichtliche Struktur des Evaluationsteams genannt. So sei lange Zeit nicht klar gewesen, wer für welche Aufgaben zuständig sei. Dies passt zu dem von einem Träger genannten Punkt, dass mit der FAU etwas mehr Kontakt wünschenswert gewesen wäre, da durch mehr Austausch auch Unklarheiten in Bezug auf die Organisation und die Zuständigkeiten hätten geklärt werden können.

Alles in allem lässt sich schlussfolgern, dass der Forderung nach Kompetenz und Glaubwürdigkeit der Evaluatoren entsprochen wurde. Durch das Evaluationsteam waren die erforderlichen Kompetenzen abgedeckt, es fanden meta-evaluative Prozesse zur Qualitätssicherung statt und es wurden den finanziellen und zeitlichen Ressourcen entsprechend angemessene Bemühungen zur Erhöhung der Glaubwürdigkeit und Akzeptanz durchgeführt.

⇨ Standard N3 wurde erfüllt.

4.2.2.1.4 Auswahl und Umfang der Informationen (N4)

> Auswahl und Umfang der erfassten Informationen sollen die Behandlung der zu untersuchenden Fragestellung zum Evaluationsgegenstand ermöglichen und gleichzeitig den Informationsbedarf des Auftraggebers und anderer Adressaten und Adressatinnen berücksichtigen.

Die Ausführungen zu diesem Standard sind in den DeGEval-Richtlinien sehr kurz gehalten. Danach sollen die potentiell zu erhebenden Informationen nach ihrer Priorität geordnet werden, indem bestimmt wird, welche Informationen für die Beantwortung der Fragestellungen zwingend erforderlich und welche Daten zwar von Interesse, für die zentralen Fragestellungen jedoch entbehrlich sind. Die Evaluatoren sollen ihre Mittel diesen Prioritäten gemäß sowie am Informationsbedarf der wichtigsten Nutzer orientiert verwenden.

Auch das JCSEE-Handbuch betont, dass die Erhebung aller zu Beginn als wichtig eingeschätzten Daten in der Regel nicht machbar und daher eine Selektion der wirklich notwendigen Informationen erforderlich ist. Dabei sollten die Bedürfnisse des Auftraggebers im Fokus stehen, es sollte jedoch auch auf weitere Bereiche geachtet werden, die z. B. für andere Beteiligte relevant sind oder sich aus der einschlägigen Literatur ergeben. Das Handbuch spricht die Empfehlung aus, den Auftraggeber aufzufordern, die potentiellen Nutzer der Evaluation ihrer Bedeutung entsprechend zu ordnen. Mit Vertretern aller Stakeholder gemeinsam sollte auch für die von ihnen benannten Themen eine Rangfolge erarbeitet werden, welche dann wiederum von Evaluierenden und Auftraggeber durchgesehen und gewichtet bzw. gekürzt werden sollte. Weiterhin sollte sich die Evaluation nicht durch zu feste

Bestimmungen bezüglich der Testverfahren einschränken lassen, indem z. B. nur standardisierte Verfahren zur Anwendung kommen.

Im vorliegenden Projekt wurden zahlreiche Daten erhoben, die letztendlich die Beantwortung der Fragestellungen umfassend ermöglichten (vgl. Abschnitt 4.1). Auch Selektionsprozesse fanden statt, um den Probanden nicht mehr Belastung zuzumuten als unbedingt notwendig. So stellte beispielsweise die Eltern-Kind-Bindung einen interessanten, aber nicht unabdingbaren Bereich dar, welcher zugunsten der Durchführbarkeit der Untersuchung sowie aufgrund der eher geringen Relevanz für die Fragestellungen schließlich nicht für die Datenerhebung ausgewählt wurde. Die Datenmenge der Evaluation war sehr umfangreich und konnte im Rahmen der Projektberichte aus Zeitgründen nicht vollumfänglich ausgewertet werden. Zwar sind weitere Analysen geplant, jedoch ist fraglich, ob diese noch in einem direkten Nutzen für die Beteiligten der Evaluation resultieren oder in erster Linie Forschungsinteressen bedienen werden. Daher kann die Anzahl der erhobenen Informationen zwar generell als angemessen gelten. Unter Berücksichtigung der aufgetretenen Verzögerungen (vgl. Standard N7) und der daraus resultierenden zu geringen zeitlichen Kapazitäten für eine umfassende Datenauswertung erscheint sie jedoch als etwas zu ambitioniert, wie auch der Auftraggeber in seiner Stellungnahme herausstellte.

Was die Erarbeitung der Erhebungsinstrumente etwas beeinträchtigte, war die Tatsache, dass die Ziele des Bundesprogramms, insbesondere auf Elternebene, zu Beginn sehr allgemein gehalten waren. Sie wurden allerdings im Rahmen von Arbeitstreffen im Verlauf der Evaluation konkretisiert. Weiterhin waren vermutete Wirkungsketten nicht expliziert worden, also auf welche Weise die Qualifizierung zum Elternbegleiter über die Fachkräfte und Eltern in Wirkungen bei den Kindern resultieren sollten. Dies erschwerte die Festlegung von Evaluationskriterien sowie die Operationalisierung der zu erwartenden Effekte und damit auch die Auswahl geeigneter Erhebungsinstrumente. Zudem ließen die Curricula eine große Varianz der Kursinhalte und -gestaltung zu, was die Evaluierbarkeit des Programms ebenfalls einschränkte.

In Bezug auf die verwendeten Erhebungsverfahren kann festgehalten werden, dass hier keine starre Ausrichtung und damit Einschränkung der Evaluation vorlag. Es wurden in allen drei Modulen

bestehende Skalen, teils in adaptierter Form, mit neu entworfenen Items verknüpft, um das Bundesprogramm möglichst vollständig abbilden und bewerten zu können. Weiterhin kamen neben quantitativen auch qualitative Verfahren zum Einsatz, um die Perspektiven der Befragten adäquat zu berücksichtigen.

Die Bedürfnisse des Auftraggebers wurden in hohem Maße berücksichtigt, hier fand hinsichtlich der Konstruktion der Erhebungsinstrumente eine kontinuierliche Absprache mit diesem sowie mit dem Kompetenzteam statt, welche auch in der Stellungnahme des Auftraggebers positiv bewertet wurde. Auch die einschlägige Literatur wurde bei der Entwicklung der Erhebungsinstrumente zu Rate gezogen, während der Einbezug weiterer Beteiligter bei diesem Schritt hingegen keine zentrale Rolle spielte. So wurden die weiteren Stakeholder, darunter die Qualifizierungsträger, aber auch Elternbegleiter und Eltern im Zuge der Pilotstudien zwar durchaus um Rückmeldung gebeten, jedoch erst, als die Erhebungsinstrumente bereits weitgehend zusammengestellt waren (mit Ausnahme eines Instruments im Rahmen von Modul 1, bei dem eine Absprache mit den Trägern stattgefunden hatte; vgl. Standard N1). Die Anregung des JCSEE-Handbuchs, eine Anordnung der verschiedenen Nutzer nach ihrer Bedeutung durch den Auftraggeber vornehmen zu lassen, wurde nicht umgesetzt. Gleiches galt für die im Handbuch empfohlene Sammlung von Themen und ihre Gewichtung durch die Beteiligten. Dieses auftraggeberorientierte Vorgehen war darauf zurückzuführen, dass der Auftraggeber den zentralen Nutzer der Evaluation darstellte.

Während die meisten Stakeholder mit ihrer Beteiligung bezüglich der Erhebungsinstrumente zufrieden waren, wie die Stellungnahmen zeigten, kritisierte der Vertreter eines Qualifizierungsträgers, in die Planung der Erhebungsinstrumente zu wenig einbezogen worden zu sein (vgl. Standard N1). Dadurch hätten die Interessen dieses Trägers nicht angemessen berücksichtigt und durch die Evaluation seine Fragen nicht ausreichend beantwortet werden können. Weiterhin wurde, wie ebenfalls bereits beim Standard N1 angegeben, der Vorschlag befürwortet, Elternbegleiter bei der Konstruktion des Elternfragebogens einzubeziehen.

In der Zusammenschau der dargelegten Daten zeigt sich, dass die Auswahl der Informationen gut gelungen war, wenngleich die Datenmenge zu umfangreich ausfiel. Vielfältige Erhebungsverfahren kamen

zur Anwendung und die potentiell interessanten Informationen wurden den Fragestellungen entsprechend selektiert. Vorhandene Literatur wurde berücksichtigt und es wurde bei der Entwicklung der Erhebungsinstrumente eng mit dem Auftraggeber zusammengearbeitet. Die weiteren Beteiligten wurden lediglich in Form von Rückmeldungen zu den bereits konstruierten Fragebögen einbezogen, was zwar überwiegend, aber nicht vollständig ihren Bedürfnissen entsprach. Dies wiederum ist angesichts der Tatsache, dass die vorliegende Evaluation einem auftraggeberorientierten Ansatz folgte, dennoch als angemessen zu bewerten. Da somit fast allen Anforderungen dieses Standards entsprochen wurde, kann dieser als überwiegend eingehalten eingestuft werden.

⇨ Standard N4 wurde überwiegend erfüllt.

4.2.2.1.5 Transparenz von Werten (N5)

Die Perspektiven und Annahmen der Beteiligten und Betroffenen, auf denen die Evaluation und die Interpretation der Ergebnisse beruhen, sollen so beschrieben werden, dass die Grundlagen der Bewertungen klar ersichtlich sind.

Laut DeGEval-Standards sind die Werte der Beteiligten und Betroffenen von großer Relevanz in verschiedenen Evaluationsphasen. Sie sollten im Rahmen der Evaluation berücksichtigt werden, um die aktive Beteiligung der Stakeholder an der Evaluation sowie ihre Nutzung der Ergebnisse zu unterstützen. Da bei der Interpretation der im Evaluationsprozess erhobenen Daten Wertvorstellungen eine zentrale Rolle spielen, sind die dabei zugrunde liegenden Werte transparent zu machen, damit die Generierung der Interpretationen nachvollziehbar wird. Während in der Kurzerklärung zu diesem Standard von einer Beschreibung der Werte der Stakeholder die Rede ist, wird nicht erläutert, in welcher Form diese erfolgen soll. Ob eine interne Klärung ausreichend ist oder eine schriftliche Darstellung in die Projektberichte integriert werden soll, wird nicht erwähnt.

Auch das JCSEE-Handbuch klärt diesen Punkt nicht auf. Dafür skizziert es einige Möglichkeiten dafür, wie die Werturteile im Rahmen der Evaluation erarbeitet werden und auf welchen Grundlagen sie basieren können. So können unter anderem die Programmziele, Gesetze

und Vorschriften sowie Leistungen einer Vergleichsgruppe für die Ergebnisinterpretation herangezogen werden. Diese kann wiederum von verschiedenen Beteiligten vorgenommen werden, darunter die Evaluatoren, der Auftraggeber, andere Stakeholder, eine Steuerungsgruppe oder eine Kombination dieser Gruppen. Zentral ist, dass der für die Erarbeitung der Werturteile angewandte Ansatz von den Evaluierenden und dem Auftraggeber sowie den weiteren Stakeholdern gemeinsam ausgewählt wird. Die Klärung der Wertvorstellungen darf allerdings nicht zu viele Ressourcen binden, damit diese nicht bei der Datenerhebung und -interpretation fehlen.

Die vorliegende Evaluation orientierte sich bei der Interpretation an den Programmzielen, wobei diese allerdings recht allgemein gefasst waren und erst im Evaluationsprozess konkretisiert wurden (vgl. Standard N4). Zudem erfolgte ein Vergleich mit einer Kontrollgruppe, welcher bereits in der Ausschreibung gewünscht worden war. Weiterhin wurden hinsichtlich der Interpretation Diskussionen mit dem Auftraggeber sowie mit zwei Steuerungsgruppen, dem Kompetenzteam und der Impulsgruppe, geführt, um einer einseitigen Perspektivenübernahme entgegenzuwirken. Zu diesem Zweck banden die Evaluatoren und der Auftraggeber auch ihre Erkenntnisse über die Werte der Qualifizierungsträger sowie der Elternbegleiter und Eltern in den Interpretationsprozess mit ein, die sie im Projektverlauf erlangt hatten.

Die Werturteile wurden somit von den Evaluierenden erarbeitet und mit dem Auftraggeber sowie den Steuerungsgruppen abgestimmt. Dieser Interpretationsansatz war zwar nicht mit allen Stakeholdern gemeinsam entwickelt worden, er wurde jedoch allen betroffenen Gruppen z. B. im Rahmen von Arbeitstreffen transparent gemacht. Nicht aktiv kommuniziert wurde er hingegen den Gruppen der Fachkräfte und Eltern, da für diese Beteiligten keine Relevanz dieser Thematik zu erwarten war. In der Tat wurden von diesen Gruppen hierzu auch keine Informationen nachgefragt.

Inwieweit dem Evaluationsteam die Berücksichtigung der Werte der verschiedenen Stakeholder gelungen war, darüber können die Stellungnahmen nur begrenzt Auskunft geben. Denn Interpretationen waren zum Zeitpunkt der Stellungnahmen lediglich im Rahmen der Kurzbroschüre veröffentlicht worden, der Abschlussbericht war noch nicht fertiggestellt und konnte daher von den Befragten nicht bewertet

werden. Auch die Kurzbroschüre war jedoch den Kursleitern beider Träger, den Eltern sowie zweien der drei Elternbegleiter nicht bekannt, sodass sie dazu keine Aussagen treffen konnten. Die übrigen Befragten kannten die Broschüre und äußerten sich durchweg positiv hinsichtlich der Berücksichtigung verschiedener Perspektiven darin. Von einem der beiden Qualifizierungsträger wurde allerdings die Thematisierung der unterschiedlichen Curricula und deren Effekte vermisst. Interessanterweise beurteilte der andere Träger die Tatsache, dass genau diese Unterscheidung nicht vorgenommen worden war, als sehr positiv.

Werden die Projektdaten und die Informationen aus den Stellungnahmen integriert, ergibt sich, dass den relevanten Beteiligten die Interpretationsgrundlagen transparent gemacht worden waren. Die Berücksichtigung ihrer Perspektiven war offenbar gelungen, wobei für eine vollumfängliche Bewertung dieses Standards interessant wäre, wie die Beteiligten den umfassenden Abschlussbericht der Evaluation diesbezüglich bewerten. Die Tatsache, dass die Erarbeitung der Bewertungsgrundlage nicht mit allen Stakeholdern gemeinsam erfolgt war, könnte negativ beurteilt werden. Jedoch ist einerseits zu bedenken, dass es sich bei der vorliegenden Evaluation um eine auftraggeberorientierte handelte, sodass z. B. der Einbezug einer Kontrollgruppe bereits von Beginn an festgelegt und auch die enge Absprache mit dem Auftraggeber sinnvoll war. Andererseits muss den Standards zufolge der Aufwand, der zur Klärung der Wertvorstellungen eingesetzt wird, gegen die vorhandenen Ressourcen abgewogen werden. Daher kann der vorliegende Standard unter den gegebenen Rahmenbedingungen und mit dem Hinweis darauf, dass die Bewertung den Abschlussbericht nicht mit einschließt und daher als vorläufig gelten muss, als eingehalten beurteilt werden.

⇨ Standard N5 wurde erfüllt.

4.2.2.1.6 Vollständigkeit und Klarheit der Berichterstattung (N6)

Evaluationsberichte sollen alle wesentlichen Informationen zur Verfügung stellen, leicht zu verstehen und nachvollziehbar sein.

Damit sie zu einer Nutzung der Evaluationsergebnisse führen, sollen Evaluationsberichte vollständig sowie klar verständlich sein, erläutern die DeGEval-Standards. Der Begriff „Evaluationsbericht" bezieht sich dabei nicht nur auf schriftliche Berichterstattung. Vielmehr ist für jede Zielgruppe eine angemessene Vermittlungsform zu wählen, wobei neben schriftlicher Darstellung auch Vorträge oder Workshops denkbar sind. Ein umfassender schriftlicher Abschlussbericht ist nicht für alle Adressaten die geeignete Form der Informationsvermittlung. Wichtig ist zudem, dass eine verständliche Sprache verwendet wird und zentrale Begrifflichkeiten definiert werden. Ein Überblick über die wesentlichen Befunde und Empfehlungen in Form einer vorangestellten Zusammenfassung oder durch Tabellen und Graphiken kann das Verständnis erleichtern.

Das JCSEE-Handbuch führt die gleichen Punkte an, geht jedoch detaillierter auf den Aspekt der Vollständigkeit ein. So sollen das Evaluandum und sein Kontext beschrieben, die Ziele der Evaluation erläutert sowie die Methodik, Ergebnisse, Schlussfolgerungen und Empfehlungen dargestellt werden. Auch sollten Fragestellungen, Ergebnisse und Probleme diskutiert und weitere Forschungsbedarfe aufgezeigt werden. Es sollen zudem verschiedene Sichtweisen berücksichtigt werden, wobei dieser Aspekt in vorliegender Arbeit beim Standard F4 näher erörtert wird. Den Bericht sollten vor der Veröffentlichung der Auftraggebende, aber auch weitere Beteiligte gegenlesen und bezüglich Klarheit, Verständlichkeit und Fairness der Darstellung bewerten. Weiterhin empfiehlt das Handbuch, Nachbereitungsveranstaltungen für die Stakeholder anzubieten. Es wird zudem darauf hingewiesen, dass die Anstrengungen, die zugunsten der Klarheit unternommen werden, nicht die Genauigkeit der Evaluation beeinträchtigen dürfen.

Bei der Evaluation von „Elternchance ist Kinderchance" waren von Beginn an ein umfassender Abschlussbericht sowie ein Zwischenbericht zur Hälfte der Projektlaufzeit geplant, wobei letzterer nicht zur Veröffentlichung gedacht war. Des Weiteren offenbarte sich im Projektverlauf der Bedarf des Auftraggebers nach einer zusätzlichen

Kurzbroschüre, in welcher die zentralen Ergebnisse der Evaluation zusammengefasst und für ein breiteres Publikum, darunter Elternbegleiter und Presse, aufbereitet wurden. Die Kurzbroschüre wurde im November 2014 veröffentlicht, der Abschlussbericht im August 2015 und der Nachbericht zur Kontrollgruppe des Moduls 3 Anfang des Jahres 2016 (vgl. Abschnitt 4.1.2). Zudem erfolgten mehrere mündliche Präsentationen im Rahmen von Arbeitstreffen mit dem Auftraggeber, dem Kompetenzteam, der Impulsgruppe, der Servicestelle, den Trägern, sowie seltener mit Kursleitern und Elternbegleitern, bei denen teilweise (Zwischen-)Ergebnisse referiert wurden.

Die mündlichen Präsentationen waren kompakt gehalten und verständlich gestaltet, wie von den Beteiligten bei den sich daran anschließenden Diskussionen rückgemeldet wurde. Auch die Kurzbroschüre war kompakt und aufgrund der angestrebten Zielgruppe nicht als fachwissenschaftliche Publikation angelegt, sondern im Hinblick auf allgemeine Verständlichkeit der Befunde formuliert. Dennoch waren ein Überblick über das Bundesprogramm und dessen Evaluation sowie eine kurze Diskussion enthalten, um auch für Personen, die das Programm bzw. die Evaluation nicht kannten, ausreichend Hintergrundinformationen bereitzustellen. Der Zwischen- und der Abschlussbericht, mit 113 bzw. 240 Seiten deutlich umfangreicher als die Kurzbroschüre mit 46 Seiten, enthielten demgegenüber ausführlichere Informationen zum Bundesprogramm, zum Evaluationsdesign und zur Methodik. Im Abschlussbericht wurden zudem die Ergebnisse und Schlussfolgerungen detaillierter geschildert und im Hinblick auf den Forschungsstand diskutiert, auch Fragestellungen für weitere Forschungsvorhaben wurden skizziert. Zusätzlich wurde dem Abschlussbericht eine Zusammenfassung vorangestellt.

Das hauptsächliche Medium zur Vorstellung der Evaluationsergebnisse stellte somit der schriftliche Bericht in Form der Kurzbroschüre und des Abschlussberichts dar, ergänzt durch eine Präsentation beim Abschluss-Workshop des Projekts, bei welchem neben Auftraggeber und Kompetenzteam auch die Impulsgruppe, die Servicestelle und Vertreter der Qualifizierungsträger anwesend waren. Dieser Workshop war überwiegend auf direkt am Projekt beteiligte Personen ausgerichtet. Der ausführliche Abschlussbericht dürfte ebenfalls in erster Linie für diese Personengruppe sowie für Wissenschaftler und Fachöffentlichkeit von Interesse sein, während mit der Kurzbroschüre ein Medium erstellt wurde, welches auch Personen erreichen soll, die nicht

speziell mit dem Projekt oder dem Fachbereich zu tun oder einfach nicht genug Zeit für den umfangreichen Abschlussbericht haben. Des Weiteren wurden von BMFSFJ, DJI und FAU Pressemitteilungen erstellt, die mittels Medienverteilern sowie durch Veröffentlichungen auf den jeweiligen Websites verbreitet wurden. Veranstaltungen zur Nachbereitung mit Beteiligten wurden mangels Ressourcen nicht durchgeführt. Bezüglich der befragten Fachkräfte und Eltern, die Interesse an Ergebnissen geäußert hatten, war ursprünglich angedacht, eine kurze Übersicht über die Evaluationsergebnisse zu erstellen und ihnen zuzusenden. Bis zum Abschluss der vorliegenden Arbeit wurde dies jedoch, ebenfalls aufgrund fehlender Mittel, nicht realisiert.

Der Abschlussbericht wurde vor der Freigabe von Auftraggeber und Kompetenzteam gelesen, welche ausführliche und hilfreiche Rückmeldungen gaben. Im Rahmen der internen Qualitätskontrolle hatte zudem ein nicht zum Evaluationsteam gehörender Wissenschaftler des DJI den Bericht gelesen. Weitere Beteiligte lasen den Abschlussbericht nicht gegen. Da er sich zum Zeitpunkt der Stellungnahmen noch in Bearbeitung befand, konnte in diesen auch nicht auf die Verständlichkeit und Fairness des Berichts eingegangen werden. Zur Kurzbroschüre waren hingegen diesbezügliche Aussagen möglich, wobei leider nicht alle Befragten die Broschüre kannten (vgl. Standard N5). Diejenigen Befragten mit Kenntnis der Broschüre schätzten deren Verständlichkeit und Fairness der Darstellung als hoch ein.

Eine weitere Frage, zu welcher die Stellungnahmen hilfreiche Informationen liefern konnten, war, inwieweit die von den Evaluatoren bereitgestellten Medien für die zu erreichenden Zielgruppen angemessen waren. Es stellte sich heraus, dass von den Gruppen, die die Kurzbroschüre kannten und die bei der Ergebnispräsentation im Rahmen des Abschluss-Workshops vertreten gewesen waren, diese Form der Vermittlung als überwiegend angemessen eingeschätzt wurde, da von Auftraggeber, Servicestelle und dem Vertreter des einen Trägers sowie der Kursleitung desselben Trägers keine anderen oder zusätzlichen Kommunikationswege gewünscht wurden. Der Vertreter des anderen Trägers befürwortete, ebenso wie der Kursleiter dieses Trägers, die Durchführung eines Workshops, bei welchem die Ergebnisse nicht nur präsentiert, sondern auch diskutiert würden und insbesondere im Hinblick auf Empfehlungen ein Austausch stattfinden könnte. Die Elternbegleiter und Eltern, welche noch gar nicht über Ergebnisse informiert gewesen waren, gaben einen schriftlichen, zusammenfassenden

Bericht als angemessene Vermittlungsform an. Ein Vertreter der Eltern sprach sich zudem für die Durchführung einer Informationsveranstaltung vor Ort aus. Der Elternbegleiter, welchem die Kurzbroschüre bekannt war, stufte diese schriftliche Berichtsform als viel Papier ein, was liegen bleibe. Diese Person empfahl, die wesentlichen Ergebnisse und Empfehlungen in Form eines prägnanten Arbeitspapiers für den Auftraggeber zusammenzufassen. Für sich selbst sowie Kollegen sei die übliche Kommunikationsform, Informationen ins Internet zu stellen, nicht angemessen, da keine Zeit vorhanden sei, um sich dort aktiv zu informieren. Hingegen könnten Informationsmails oder Newsletter per E-Mail ein geeignetes Medium darstellen.

Zusammengenommen ergibt sich für den vorliegenden Standard, dass die Einhaltung der Kriterien der Klarheit und Vollständigkeit des Berichts durch die Kurzbroschüre und den Abschlussbericht in unterschiedlichem Ausmaß gegeben war, wobei sich beide Publikationsformen sinnvoll ergänzten. So war die Kurzbroschüre knapp gehalten und eher auf Klarheit ausgelegt und wurde auch als verständlich und fair bewertet. Der Abschlussbericht stellte hingegen das vollständigere Medium dar, in welchem eine wissenschaftlichere Ausdrucksweise zur Anwendung kam, welche seiner Zielgruppe angemessen war. Inwieweit der Abschlussbericht von anderen Beteiligten als verständlich, vollständig und fair eingeschätzt wird, kann an dieser Stelle nicht beantwortet werden, da er nicht von verschiedenen Beteiligten gegengelesen und bewertet wurde. Die gewählten Darstellungsformen für die Vermittlung der Evaluationsergebnisse passten nur teilweise zu den Bedürfnissen der Beteiligten. Nachbereitungsworkshops, die von einigen Befragten gewünscht wurden, wurden aufgrund mangelnder Ressourcen nicht realisiert. Dies führt zu Abzügen bei der Bewertung des Standards, sodass dieser als teilweise eingehalten bewertet wird.

⇨ Standard N6 wurde teilweise erfüllt.

4.2.2.1.7 Rechtzeitigkeit der Evaluation (N7)

> Evaluationsvorhaben sollen so rechtzeitig begonnen und abgeschlossen werden, dass ihre Ergebnisse in anstehenden Entscheidungsprozessen bzw. Verbesserungsprozessen einfließen können.

Den DeGEval-Standards zufolge soll eine Evaluation von Beginn an so geplant und durchgeführt werden, dass sie ihre Ergebnisse zu dem Zeitpunkt liefern kann, zu dem diese benötigt werden. Dies betrifft eine rechtzeitige Vorbereitung, Ausschreibung und Beauftragung ebenso wie eine realistische Zeitplanung für die Durchführung der Evaluation, welche koordinative Aufgaben, die Vorbereitung der Datenerhebungen, Fehlerprüfungen sowie ausreichend Zeit für Analyse und Interpretation der Daten berücksichtigen und zudem Pufferzeiten umfassen soll. Die Evaluation soll den vorgesehenen Nutzern Zwischenergebnisse und Schlussberichte so frühzeitig vorlegen, dass diese sie verwenden können. Damit eine Evaluation Wirkungen nach sich ziehen kann, ist es erforderlich, den Evaluationsprozess zeitlich an bevorstehenden Entscheidungen bzw. Optimierungen auszurichten. Auch hier ist zwischen der Abgabe eines Evaluationsberichts und dem Treffen einer Entscheidung bzw. dem Einleiten von Verbesserungen zusätzlich Zeit einzuplanen, sodass die Ergebnisse entsprechend rechtzeitig zu erbringen sind. Auch vorläufige Befunde und Zwischenergebnisse werden häufig gewünscht und sind insbesondere dann erforderlich, wenn diese für aktuelle Handlungsschritte der Nutzer relevant sind. Die dazu benötigten Ressourcen einschließlich eines Zeitkontingents müssen ebenfalls eingeplant werden.

Der entsprechende Standard im JCSEE-Handbuch enthält vergleichbare Erläuterungen. Zusätzlich wird genauer darauf eingegangen, wer zu den vorgesehenen Nutzern zu zählen ist. Diese umfassen Personen, die einen Anspruch auf Informationen zu den Evaluationsergebnissen haben, weil sie die Evaluation in Auftrag gegeben haben, die rechtliche Verantwortung für das Evaluandum tragen, die evaluierte Maßnahme mit Spenden, Zeit oder Arbeit unterstützt oder für die Evaluation wichtige Informationen geliefert haben. Auch weitere Beteiligte und Betroffene, z. B. Programmplaner oder Medienvertreter können zur Gruppe der vorgesehenen Nutzer gehören. Evaluatoren und Auftraggeber müssen gemeinsam alle vorgesehenen Nutzer identifizieren,

kontaktieren sowie informieren, und zwar in einem für diese jeweils geeigneten Berichtsmodus (vgl. Standard N6). Die Evaluatoren sollten die Bedarfe des Auftraggebers sowie der weiteren Nutzer hinsichtlich der Ergebnisse, der geeigneten Berichtsform und des passenden Zeitpunkts herausfinden. Die Verantwortung für die Verbreitung der Informationen sollten Evaluierende und Auftraggeber sich teilen und zur Vermeidung von Zuständigkeitskonflikten in Form eines Verbreitungsplans schriftlich festhalten. Auch die Freigabe von Berichten und die redaktionelle Verantwortlichkeit sollten schriftlich fixiert werden. Eine solch offizielle Übereinkunft stärkt auch die Evaluatoren in ihrem Recht bzw. ihrer Pflicht, neben dem Auftraggeber auch andere Nutzer angemessen zu informieren. Weiterhin weist das JCSEE-Handbuch darauf hin, dass keine ungenauen oder lückenhaften Informationen vermittelt werden dürfen, um Abgabetermine von Berichten einzuhalten.

Bei der hier dargestellten Evaluation war in der Auftragsbeschreibung im Rahmen des Vergabeverfahrens ein Zeitplan enthalten, welcher den Beginn der Evaluation im März 2012 und den Abschluss im Dezember 2014 vorsah. Zudem wurden darin ein Zwischenbericht Anfang 2013 sowie ein Abschlussbericht zum Ende der Projektlaufzeit gefordert. Im Werkvertrag wurden diese Berichtspflichten konkretisiert, indem für den Zwischenbericht eine Abgabe zum 01.04.2013 und für den Abschlussbericht als Abgabetermin der 31.12.2014 vereinbart wurde. Die Evaluation war grundlegend so geplant, dass diese Fristen eingehalten werden konnten, wobei jedoch kaum Pufferzeiten anberaumt waren.

Der Zwischenbericht konnte zum vereinbarten Zeitpunkt abgegeben werden. Die Abgabe des Abschlussberichts hingegen verzögerte sich zunächst bis Januar 2015, die endgültige Fassung wurde im Juni 2015 fertiggestellt und im August 2015 veröffentlicht. Diese Verzögerung war unter anderem darauf zurückzuführen, dass die sehr umfangreichen Datenerhebungen mehr Zeit in Anspruch nahmen als geplant. Dies lag teilweise in zum Beginn der Evaluation bzw. zum Zeitpunkt der Angebotserstellung nicht absehbaren Komplikationen begründet. So mussten bei den Online-Befragungen im Modul 1 aus Datenschutzgründen zwei unterschiedliche Zugänge zu den Qualifizierungsteilnehmern beider Träger gewählt werden, was mit einer Verdoppelung des dafür eingeplanten Zeitaufwands einherging (vgl. Standard F2). Außerdem waren diese Befragungen zeitlich an die Termine der

Qualifizierungskurse gebunden und damit nicht flexibel durchführbar. Auch die Erarbeitung der Fragebögen gestaltete sich aufwendiger als gedacht, da die Curricula den Kursleitern viele Freiheiten ließen und deshalb mehr Arbeitstreffen mit Programmverantwortlichen notwendig waren, um die Umsetzung der Qualifizierungen operationalisieren zu können. Weiterhin war in der Kontrollgruppe ein sehr geringer Rücklauf zu verzeichnen (vgl. Abschnitt 4.1.3.2), sodass deren Rekrutierung deutlich mehr Ressourcen band als erwartet. Da zudem im Modul 3 die Befragungen denen des Moduls 1 zeitlich nachgeschaltet waren und deutlich länger andauerten als geplant, konnten die Ergebnisse mit Einschluss der Kontrollgruppe für dieses Modul erst verspätet im Juni 2015 berichtet werden. Die Veröffentlichung dieses Berichts erfolgte Anfang des Jahres 2016.

Insgesamt muss festgehalten werden, dass die Umsetzung des Evaluationsdesigns die vorhandenen personellen und zeitlichen Ressourcen deutlich überstieg. Lediglich durch den persönlichen Einsatz der Mitglieder des Evaluationsteams in Form von erheblicher Mehrarbeit sowie durch gewisse Abstriche bei den Datenanalysen konnte die Erbringung der vertraglich vereinbarten Leistungen gesichert werden. Hierzu ist zu erwähnen, dass bei den Datenauswertungen die geplanten Mehrebenenanalysen nicht realisiert werden konnten, welche jedoch im Rahmen der Dissertation einer Projektmitarbeiterin nachgeliefert werden. Zudem konnten für den Abschlussbericht die erhobenen quantitativen und qualitativen Daten nicht in der Tiefe ausgewertet werden, die dem hohen Aufwand für die Datenerhebung angemessen gewesen wäre.

Bei der Bewältigung der Evaluationsaufgaben spielte zudem die bereits erwähnte hohe Personalfluktuation im Evaluationsteam eine Rolle, welche neben einem erhöhten Einarbeitungsaufwand auch Erfahrungsverlust mit sich brachte (vgl. Abschnitte 4.1.2 und 4.2.2.1.3).

Einen weiteren Grund für die verzögerte Abgabe des Abschlussberichts stellten redaktionelle Überarbeitungen desselben dar, die im Zeitraum von Januar 2015 bis Juni 2015 vorgenommen wurden. Auftraggeber und Kompetenzteam gaben den Evaluatoren viele Rückmeldungen, die sich über mehrere Korrekturschleifen erstreckten und als zeitaufwendig erwiesen.

Anteil an der verspäteten Abgabe des Abschlussberichts hatten zudem die Auswertungen zum Vergleich der beiden Curricula (vgl. Standard N2), sowie die Kurzbroschüre, deren Erstellung der Auftraggeber

erst relativ spät wünschte (vgl. Abschnitt 4.1.2). Diese beiden Aufgaben banden zusätzliche Ressourcen des Evaluationsteams. Um dies zu kompensieren, wurde gemeinsam mit dem Auftraggeber beschlossen, auf manche Auswertungen zu verzichten und einige Erhebungen in geringerem Umfang durchzuführen als ursprünglich geplant. Zudem wurde vereinbart, die Abgabefrist des Abschlussberichts zu verlängern. Als problematisch erwies sich dabei die Tatsache, dass die Arbeiten am Abschlussbericht bis in das Jahr 2015 und somit über die Vertragslaufzeiten der Projektmitarbeiter hinaus andauerten.

Bei der Bewertung der Rechtzeitigkeit der Evaluation ist zu bedenken, dass eine Evaluation Ergebnisse zu dem Zeitpunkt liefern sollte, zu dem diese benötigt werden. Dazu ist sie entsprechend zu planen und durchzuführen. Im Hinblick auf den formativen Teil der Evaluation erscheint dies gelungen, da im Verlauf Optimierungsmöglichkeiten aufgezeigt werden konnten, was zu Nachsteuerungsprozessen führte. In Bezug auf die summative Evaluation ist allerdings festzuhalten, dass deren Ergebnisse der ursprünglichen Ausschreibung zufolge erst Ende 2014 hätten benötigt werden sollen, da zu diesem Zeitpunkt die Abgabe des Abschlussberichts und das Ende der Evaluation geplant waren. Auch das Bundesprogramm selbst hatte eine geplante Laufzeit bis Ende 2014, sodass das Ende des Programms mit dem Ende der Evaluation zusammenfiel. Im Evaluationsverlauf zeigte sich allerdings, dass die Ergebnisse der summativen Evaluation doch früher benötigt wurden, um bei den Planungen der nächsten Programmphase (Elternchance II) Verwendung zu finden (vgl. Abschnitt 4.1.2). Dies betraf die vergleichenden Auswertungen zu den beiden Curricula, die der Auftraggeber bereits im Frühjahr 2014 wünschte. Hierbei ist zum einen als problematisch zu bewerten, dass diese Auswertungen auf vorläufigen Daten basierten. Denn wenngleich die Stichprobenzahlen bereits einen hohen Umfang erreicht hatten, waren zu diesem Zeitpunkt noch nicht alle Erhebungen abgeschlossen. Zum anderen ist fraglich, ob die Evaluation nicht von Beginn an anders hätte angelegt werden können, ob also nicht absehbar war, dass die Ergebnisse in einen Überarbeitungsprozess einfließen könnten, der deutlich vor dem Ende des Programms stattfinden würde.

Die von den Standards geforderte Identifizierung vorgesehener Nutzer und deren Berichtsbedarfe gemeinsam mit dem Auftraggeber war dahingehend erfolgt, dass im Verlauf der Evaluation über die Adressaten gesprochen wurde sowie Vor- und Nachteile verschiedener

Publikationsformate durch das Evaluationsteam aufgezeigt wurden. Die redaktionelle Verantwortlichkeit war im Werkvertrag geregelt, ebenso die Freigabe von Berichten. Hinsichtlich der Zuständigkeit für die Verbreitung der Ergebnisse ist festzuhalten, dass der Auftraggeber das ausschließliche Nutzungsrecht hatte, während den Evaluatoren ein einfaches Nutzungsrecht für Lehre und Forschung nach vorheriger Abstimmung mit dem Auftraggeber eingeräumt wurde. Die *Rechte* waren somit klar geregelt, die *Pflicht* der Ergebnisverbreitung war hingegen nicht festgelegt. Welche Maßnahmen zur Verbreitung der Evaluationsergebnisse ergriffen wurden und wie diese zu bewerten sind, wird beim folgenden Standard N8 diskutiert. Wie der Auftraggeber in seiner zweiten Stellungnahme anmerkte, sei die Tatsache, dass die Ergebnisverbreitung nur teilweise festgelegt war, als Merkmal der Auftragsforschung zu sehen. Demnach stelle zunächst der Auftraggeber den Nutzer der Evaluationsergebnisse dar und dieser entscheide über die weitere Verbreitung.

Dass die Rechtzeitigkeit der Ergebnisvermittlung nicht für alle Stakeholder erbracht worden war, wurde durch die Stellungnahmen deutlich. So waren der Auftraggeber, die Servicestelle und ein Träger mit den Informationen, die sie über Ergebnisse erhalten hatten, zufrieden. Auch der Kursleiter dieses Trägers bewertete die Ergebnisvermittlung als rechtzeitig, da ohnehin wenig Zeit zur Verfügung gestanden habe und die in dieser Zeit gewonnenen Erkenntnisse bereits sehr umfangreich seien. Zwischenergebnisse wären zwar interessant gewesen, jedoch würden diese in der Regel auf vorläufigen Daten basieren und seien damit ein Stück weit angreifbar. Verständnis dafür, dass nicht früher Ergebnisse mitgeteilt worden waren, äußerte auch der Vertreter des anderen Trägers, betonte aber dennoch, dass mehr Austausch über Zwischenstände sinnvoll gewesen wäre. Der Kursleiter dieses Trägers kritisierte ebenfalls, über Ergebnisse weder im Verlauf noch gegen Ende der Evaluation in Kenntnis gesetzt worden zu sein. Da er vergleichsweise viel Kontakt zu den zuständigen Mitarbeitern seines Trägers habe und dennoch diesbezüglich keine Informationen erhalten habe, gelte dies wohl ebenso für andere Kollegen. Er erläuterte weiterhin, dass bei rechtzeitiger Information, beispielsweise Mitte oder Ende des Jahres 2013, die daraus abgeleiteten Empfehlungen in den Qualifizierungen hätten umgesetzt werden können, was aber nun nicht geschehen sei. Zudem sei nicht bekannt, dass die Evaluation zu

Änderungen für den zweiten Teil des Bundesprogramms geführt habe, sodass der Nutzen der Evaluation fraglich sei. Interessanterweise flossen die Evaluationsergebnisse sehr wohl in die Überarbeitung der Modulhandbücher für die zweite Phase von „Elternchance ist Kinderchance“ ein, was auch den Trägern bekannt war, wie ihren Stellungnahmen entnommen werden konnte. Den Kursleiter hingegen hatte diese Information offensichtlich nicht erreicht.

Von den übrigen Befragten hatten die Vertreter der Elterngruppe bislang gar keine Informationen über Ergebnisse erhalten, wünschten sich dies jedoch. Ebenso verhielt es sich bei den beiden Elternbegleitern, welche nicht am 2. Bundeskongress im November 2014 teilgenommen hatten. Der Elternbegleiter mit Teilnahme an diesem Kongress war dort über Ergebnisse informiert worden und hatte in diesem Zusammenhang auch die Kurzbroschüre erhalten. Er äußerte sich sehr positiv darüber und bemerkte hinsichtlich Zwischenergebnissen, dass es hierzu keine Informationen gegeben habe, dass jedoch bei ihm und vermutlich auch bei den Kollegen gar keine zeitlichen Kapazitäten für solche Zwischenstände vorhanden seien.

Die hier dargestellte Kritik an der Ergebnisverbreitung wird im folgenden Standard N8 aufgegriffen und auf die Nutzung der Evaluation bezogen.

Zusammenfassend ist festzuhalten, dass in der vorliegenden Evaluation einige Komponenten dieses Standards nicht erfüllt werden konnten. Die grundlegende zeitliche Planung hätte eigentlich eine Einhaltung der Berichtstermine ermöglicht, die Datenerhebungen stellten sich jedoch als im Verhältnis zu den gegebenen Ressourcen zu umfangreich und aufwendig heraus. In der Folge konnte der Abschlussbericht der Evaluation erst verspätet abgegeben werden, die Kontrollgruppe des Moduls 3 musste ausgekoppelt und in einem gesonderten Bericht nachgeliefert werden. Während hinsichtlich der Berichte somit Verzögerungen gegenüber dem ursprünglichen Zeitplan zu verzeichnen waren, konnte die Evaluation dem Auftraggeber als zentralem Nutzer dennoch die gewünschten Ergebnisse zum erforderlichen Zeitpunkt liefern und damit ihre Zielsetzung erfüllen. Dies betraf sowohl die formative als auch die summative Ausrichtung der Evaluation. Kritisch zu betrachten ist allerdings, dass die für die Fortführung des Bundesprogramms relevanten Auswertungen (Vergleiche der Curricula) bereits im Frühjahr 2014 benötigt wurden, während die ursprüngliche

Anlage der Evaluation deren Ende und damit das Vorliegen der summativen Ergebnisse erst zum Ende des Jahres 2014 vorgesehen hatte. In der Folge konnten zu dem Zeitpunkt, zu dem die Evaluationsergebnisse für einen Überarbeitungsprozess benötigt wurden, diese lediglich auf der Basis vorläufiger Daten geliefert werden, was die Rechtzeitigkeit einschränkt. Die übrigen Berichte umfassten jedoch die vollständigen Daten sowie genaue Auswertungen, wenngleich aus zeitlichen Gründen nicht alle geplanten Analysen umgesetzt werden konnten. Die Rechtzeitigkeit der Ergebnisvermittlung war nicht für alle Stakeholder gelungen, was sich in dem Wunsch vieler Stellungnehmer nach mehr Austausch und Informationen, auch über Zwischenergebnisse, äußerte. Da insgesamt einige Anforderungen dieses Standards aus verschiedenen Gründen nicht eingehalten wurden, ergibt sich folgende Bewertung:

⇨ Standard N7 wurde teilweise erfüllt.

4.2.2.1.8 Nutzung und Nutzen der Evaluation (N8)

> Planung, Durchführung und Berichterstattung einer Evaluation sollen die Beteiligten und Betroffenen dazu ermuntern, die Evaluation aufmerksam zur Kenntnis zu nehmen und ihre Ergebnisse zu nutzen.

Damit eine Evaluation wie vorgesehen genutzt wird, soll sie laut DeGEval-Standards den Kriterien der Durchführbarkeit, Fairness und Genauigkeit entsprechen und zudem die Anforderungen der Standards N1 bis N7 erfüllen. Es wird weiterhin erwähnt, dass die Nutzung der Evaluation durch eine angemessene Beteiligung der Adressaten in der Planungs- und Durchführungsphase sowie bei der Aufbereitung der Ergebnisse unterstützt wird. Die Stakeholder sollen die Erwartung haben, dass sie einen Nutzen aus der Evaluation ziehen können. Auch regelmäßige, transparente Rückmeldungen im Evaluationsverlauf werden empfohlen, da dann neben dem Ergebnisnutzen auch ein Prozessnutzen eintreten kann.

Im JCSEE-Handbuch wird spezifiziert, wie die Unterstützung der Beteiligten durch die Evaluierenden realisiert werden kann. Evaluatoren sollen demnach nicht erwarten, dass die Erstellung und Abgabe eines Evaluationsberichts automatisch zur Nutzung der Ergebnisse

führt. Der mögliche Nutzen der Evaluation soll den verschiedenen Beteiligten bereits zu Beginn aufgezeigt werden, sie sollen in die Planung und gegebenenfalls Durchführung einbezogen werden und die Berichterstattung soll offen, aufrichtig und konkret sein. Die Evaluierenden sollten für mögliche Rückfragen ansprechbar sein und den Stakeholdern auch nach Freigabe des Abschlussberichts bei der Interpretation und Umsetzung der Ergebnisse helfen, sofern die zeitlichen, finanziellen und personellen Ressourcen dies gestatten. Sie sollten jedoch nicht die Rolle des Auftraggebers übernehmen, indem sie z. B. aus den Ergebnissen abgeleitete Empfehlungen selbst umsetzen.

Bei der Bewertung des Standards N8 ist zu beachten, dass er sich auf die Kriterien der Durchführbarkeit, Fairness und Genauigkeit sowie auf die Standards N1 bis N7 bezieht und somit eine Art „Gesamt-Standard" darstellt. Daher wird an dieser Stelle auch auf die Bewertungen der übrigen Standards zurückgegriffen. Betrachtet man diese in der Zusammenschau (vgl. die Übersicht in Abschnitt 4.2.3), so wird deutlich, dass bei der vorliegenden Evaluation die Mehrheit der Standards erfüllt oder überwiegend erfüllt wurde. Nur wenige wurden teilweise erfüllt und kein Standard wurde als nicht eingehalten bewertet. Zu beachten ist allerdings, dass die Rechtzeitigkeit der Evaluation als nur teilweise erfüllt bewertet wurde, welche einen Aspekt darstellt, dem ein direkter Einfluss auf die Möglichkeit der Evaluationsnutzung zugesprochen werden kann (vgl. Standard N7). Auch die Abstriche hinsichtlich der Passung zwischen den gewählten Berichtsformen und den Bedarfen der Beteiligten, die beim Standard N6 aufgedeckt wurden, sind als relativ gravierend einzustufen, da Ergebnisse stets zuerst vermittelt und rezipiert werden müssen, bevor sie genutzt werden können.

Wie in Standard N1 erörtert, waren die Identifizierung der Stakeholder und ihre Beteiligung bei Planung und Durchführung der Evaluation überwiegend gelungen. Bei der Aufbereitung der Ergebnisse war nur der Auftraggeber einbezogen worden, ein Umstand, der jedoch lediglich von dem Vertreter eines Trägers beanstandet wurde (vgl. Standard N1) und zudem als angemessen zu bewerten ist, da es sich um eine auftraggeberorientierte Evaluation handelte. Häufiger kritisiert wurde in den Stellungnahmen hingegen die Tatsache, dass es zu wenig oder keine Informationen bezüglich Ergebnissen bzw. Zwischenergebnissen im Projektverlauf gegeben hatte (vgl. Standard N7). Ob-

wohl einige Stakeholder mit der Ergebnisverbreitung zufrieden waren, sind dennoch die kritischen Anmerkungen ernst zu nehmen, da ohne eine Kenntnis der Ergebnisse diese selbstverständlich auch nicht genutzt werden können.

Es ist allerdings anzumerken, dass der Abschlussbericht sich noch in Bearbeitung befand, als die Stellungnahmen erhoben wurden (vgl. Standard N6), sodass dessen Verbreitung schlicht noch nicht möglich gewesen war. Hingegen war zu diesem Zeitpunkt die Kurzbroschüre bereits einige Zeit veröffentlicht und dennoch nur einem Teil der befragten Stakeholder bekannt gewesen. Die Kurzbroschüre war als gedruckte Version beim 2. Bundeskongress Elternbegleitung im November 2014 verteilt worden, ebenso bei der wissenschaftlichen Abschlusstagung im Dezember 2014. Ansonsten war sie als Internetpublikation über die Website des Bundesprogramms kostenlos erhältlich. Jedoch war, abgesehen von den bereits erwähnten Pressemitteilungen (vgl. Standard N6) und einer Rundmail des Auftraggebers, lediglich auf dieser Website sowie auf den Homepages des BMFSFJ und des DJI ein Hinweis auf die Broschüre erschienen, auf denen potentielle Leser sich selbst aktiv hätten umsehen müssen. Wie der Kursleiter eines Trägers sowie ein Elternbegleiter bemerkten, sei dies nicht der geeignete Weg, um Neuigkeiten zu verbreiten, da man sich eben nicht regelmäßig erkundige, ob es etwas Neues gebe. Beide schlugen beispielsweise halbjährliche Newsletter als Vermittlungsweg auch für Zwischenergebnisse vor, wobei der Kursleiter ansprach, dass für deren Erstellung natürlich auch Kapazitäten vorhanden sein müssten. Er sprach sich zudem für die Durchführung eines Workshops zur Hälfte des Programms mit dem Evaluationsteam zusammen aus.

Es sind somit gewisse Mängel in der Ergebnisverbreitung hinsichtlich der Kurzbroschüre festzustellen, da die Ergebnisse nicht alle Stakeholder und damit potentiellen Nutzer erreicht hatten. Zwar sind nach der Freigabe des Abschlussberichts Vorträge und weitere Publikationen durch das Evaluationsteam geplant, diese dürften jedoch eher die interessierte Fachöffentlichkeit ansprechen und weniger die Ergebnisnutzung durch die Beteiligten und Betroffenen unterstützen. Auch hier sei allerdings die Einordnung dieser Evaluation als Auftragsforschung in Erinnerung gerufen. Wie sowohl der Auftraggeber als auch das Mitglied des Evaluationsteams in ihrer jeweils zweiten Stellungnahme anmerkten, stellte das BMFSFJ den zentralen Nutzer der Ergebnisse dar, der über deren weitere Verbreitung entschied. Es sei

bei vergleichbaren Aufträgen üblich, dass der Auftraggeber die Ergebnisse veröffentliche. Im vorliegenden Projekt geschah dies in Form der Kurzbroschüre, deren Verbreitung wie oben beschrieben vorgenommen wurde. Positiv hervorzuheben sind zudem die beiden Bundeskongresse Elternbegleitung, die das BMFSFJ veranstaltete. Dazu waren neben Elternbegleitern und Fachkräften aus *Elternbegleitung Plus*-Einrichtungen auch Vertreter des Evaluationsteams, des Kompetenzteams und weiterer Stakeholder eingeladen und es wurden Erfahrungen und Erkenntnisse ausgetauscht sowie insbesondere beim zweiten Bundeskongress auch Evaluationsergebnisse vermittelt.

Wie die Stellungnahmen weiterhin ergaben, war dem Bedarf einiger Stakeholder nach Zwischenergebnissen nicht ausreichend nachgekommen worden (vgl. Standard N7). Es hatte im Projektverlauf zwar auch einen Zwischenbericht gegeben, der jedoch nur für den Auftraggeber bestimmt und nicht zur Publikation vorgesehen war. Dass dennoch einige Beteiligte offenbar mit mehr Zwischenergebnissen gerechnet hatten, könnte auf einen diesbezüglichen Mangel an Kommunikation hinweisen.

Eine Aufklärung der Beteiligten darüber, welchen Nutzen die Ergebnisse für sie haben könnten, fand zwar frühzeitig im Rahmen von Arbeitstreffen mit Auftraggeber, Servicestelle und Trägern statt, und auch den Fachkräften und Eltern wurde im Zusammenhang mit ihrer Befragung der Zweck und mögliche Nutzen der Evaluation aufgezeigt. Allerdings war der Nutzen zunächst noch wenig spezifiziert und konnte nur in allgemeiner Form aufgezeigt und erst im Lauf der Evaluation konkretisiert werden. Der Anforderung eines offenen, aufrichtigen und konkreten Berichts wurde entsprochen, wie an anderer Stelle näher erläutert wird (vgl. Standards F4 und F5). Des Weiteren standen die Evaluatoren für Rückfragen zu den Ergebnissen und Berichten zur Verfügung. Eine aktive Unterstützung der Beteiligten bei der Ergebnisnutzung nach der Erstellung und Freigabe des Abschlussberichts konnte jedoch nicht realisiert werden, da solche Aktivitäten durch die Evaluationsressourcen nicht abgedeckt und auch nicht geplant gewesen waren. Dass seitens einiger Beteiligter ein Bedarf an Nachbereitung bestand, z. B. in Form eines Workshops mit Diskussion der Ergebnisse, wurde im Rahmen der Stellungnahmen deutlich (vgl. Standard N6). Schließlich ist festzuhalten, dass die Evaluierenden nicht die Rolle des Auftraggebers einnahmen, da sie die Umsetzung von Empfehlungen

nicht selbst übernahmen, sondern z. B. die Überarbeitung der Curricula durch den Auftraggeber und die Träger realisiert wurde.

Zusammengenommen ergibt sich, dass einigen Anforderungen dieses Standards entsprochen werden konnte. Andere hingegen, wie z. B. die Rechtzeitigkeit und Passung der Berichtsform oder die Unterstützung bei der Ergebnisnutzung auch nach der Fertigstellung des Abschlussberichts, konnten im vorliegenden Evaluationsprojekt nicht erfüllt werden. Für den vorliegenden Standard ergibt sich damit folgende Bewertung:

⇨ Standard N8 wurde teilweise erfüllt.

4.2.2.2 Durchführbarkeitsstandards

Die Durchführbarkeitsstandards sollen sicherstellen, dass eine Evaluation realistisch, gut durchdacht, diplomatisch und kostenbewusst geplant und ausgeführt wird.

4.2.2.2.1 Angemessene Verfahren (D1)

> Evaluationsverfahren, einschließlich der Verfahren zur Beschaffung notwendiger Informationen, sollen so gewählt werden, dass Belastungen des Evaluationsgegenstandes bzw. der Beteiligten und Betroffenen in einem angemessenen Verhältnis zum erwarteten Nutzen der Evaluation stehen.

Die DeGEval-Standards weisen darauf hin, dass die Verfahren, die im Rahmen einer Evaluation zum Einsatz kommen, einerseits wissenschaftliche Gütekriterien erfüllen sollen, andererseits jedoch das Evaluandum und die Beteiligten nicht mehr als unbedingt notwendig belasten oder stören dürfen. Häufig können die wissenschaftlich anerkanntesten Verfahren nicht zur Anwendung kommen, weil sie sich als in zeitlicher oder finanzieller Hinsicht zu aufwendig oder als im gegebenen Kontext ethisch nicht akzeptabel erweisen. Die Evaluatoren haben die Aufgabe, die Vor- und Nachteile sowie die Aussagekraft der ausgewählten Verfahren kenntlich zu machen und zu begründen.

Die Autoren des JCSEE-Handbuchs erwähnen zudem, dass das Evaluationspersonal zur Anwendung der Verfahren ausreichend qualifiziert sein muss und gegebenenfalls Schulungen notwendig sein können, um dies zu erreichen. Die Verfahren sollen einen vertretbaren Aufwand erfordern und der Verfügbarkeit der zu Befragenden entsprechen. Es sollten Alternativpläne entwickelt und zeitliche sowie finanzielle Puffer eingeplant werden, um eventuellen Schwierigkeiten begegnen zu können. Der Zeitplan sollte mit den Auftraggebenden gemeinsam auf Machbarkeit geprüft werden. Ebenso sollte eine Klärung der Durchführbarkeit und des praktischen Nutzens der Erhebungsverfahren anhand einer Pilotstudie erfolgen.

Bei der vorliegenden Evaluation waren die Kompetenzen zur Durchführung der Verfahren im Evaluationsteam entweder bereits vorhanden oder wurden vor der Datenerhebung durch Weiterbildungen erzielt. So ließen sich die Mitglieder der FAU intensiv im Bereich der qualitativen Methoden schulen, in dem sie bis dahin wenig Erfahrung hatten. Die Evaluierenden hatten die Vor- und Nachteile sowie die Aussagekraft der gewählten Verfahren dem Auftraggeber verdeutlicht und diese auch bei Arbeitstreffen mit den Qualifizierungsträgern diskutiert. Insbesondere mit dem Auftraggeber fand diesbezüglich sowie hinsichtlich des Zeitplans eine enge Absprache statt. Auch in der Stellungnahme bezeichnete der Auftraggeber das Evaluationsdesign und die Wahl der Erhebungsinstrumente als plausibel und hob den engen Abstimmungsprozess positiv hervor. Allerdings waren keine Alternativpläne entwickelt und auch kaum zeitliche oder finanzielle Puffer eingeplant worden (vgl. Standard N7). Jedoch wurde eine Pilotstudie durchgeführt, um die Verständlichkeit und Durchführbarkeit der Verfahren zu testen.

Wie in den Standards angegeben, musste die Genauigkeit der Datenerhebung gegen ihre Durchführbarkeit abgewogen werden. Ein sinnvoller Kompromiss wurde dadurch erzielt, dass wissenschaftlich anerkannte, valide Fragebögen in Kombination mit eigens entwickelten Items verwendet wurden und von den validen Instrumenten Kurzformen bzw. adaptierte Versionen zur Anwendung kamen, um den Aufwand für die Teilnehmenden gering zu halten (vgl. Standard G5). Zu diesem Zweck wurden auch die Daten aus dem Prätest analysiert und solche Items, die dabei kaum Varianz aufwiesen, nicht in den Posttest-Fragebogen aufgenommen. Zudem wurden nur Informationen

erhoben, die für die Evaluation notwendig waren (vgl. Standard N4). Auch die quantitativen und qualitativen Erhebungen fanden in Ergänzung zueinander statt, d. h. dass in den qualitativen Interviews keine Informationen erfragt wurden, die bereits in der quantitativen Befragung erhoben worden waren.

Ein weiterer Kompromiss wurde beim Zugang zur Elternstichprobe eingegangen. Dem ursprünglichen Plan zufolge sollten Fachkräfte, die an der Online-Befragung des DJI teilgenommen hatten, kontaktiert werden und in der Folge Eltern für die Teilnahme an der FAU-Befragung vermitteln. Aus diesen vermittelten Eltern sollte eine Zufallsstichprobe für die tatsächliche Teilnahme an der Befragung gezogen werden. Der Rücklauf der vermittelten Eltern stellte sich jedoch als deutlich geringer als geplant heraus und hätte nur durch beträchtlichen Mehraufwand verbessert werden können. So hätten die Evaluatoren durch Besuche vor Ort die Teilnahmebereitschaft möglicherweise erhöhen können, was jedoch seitens des Evaluationsteams mit immensen zeitlichen und finanziellen Kosten verbunden gewesen wäre und auch für die Fachkräfte deutlich mehr Aufwand bedeutet hätte, da sie dann Treffen zwischen Evaluierenden und potentiellen Teilnehmenden hätten arrangieren müssen. Daher wurde die Idee einer Zufallsstichprobe zugunsten einer Rekrutierungsstrategie aufgegeben, bei welcher alle von den Fachkräften vermittelten Eltern in die Befragung mit aufgenommen wurden. Daraus ergaben sich Abstriche bei der Genauigkeit aufgrund der Selektivität der Elternstichprobe, welche jedoch durch die weitaus höhere Praktikabilität dieses Vorgehens ausgeglichen wurden (vgl. hierzu auch Standard G5).

Bei der Datenerhebung wurde in hohem Maße Rücksicht auf die Verfügbarkeit der Teilnehmenden genommen. So wurde zur Befragung der Fachkräfte ein Online-Fragebogen eingesetzt, welcher von diesen zeitlich flexibel ausgefüllt werden konnte. Bei der telefonischen Befragung der Eltern passten sich die Interviewer stets an deren Tagesablauf an und führten das Interview zu der von den Befragten gewünschten Zeit. Und auch bei den persönlichen Gesprächen im Rahmen z. B. der Fokusgruppen im Modul 2 der Evaluation wurde der Aufwand für die Befragten möglichst gering gehalten, indem die Evaluatoren die Gespräche vor Ort durchführten und dazu selbst anreisten.

Dass die angewendeten Verfahren so gut wie keine unnötigen Belastungen für die Befragten bedeuteten, lässt sich auch aus den Stellungnahmen ableiten. So bewerteten alle Elternbegleiter und Eltern die

Befragungen, an denen sie teilgenommen hatten, sowohl hinsichtlich der Verständlichkeit als auch in Bezug auf die Dauer positiv. Die Eltern hoben zudem die Freundlichkeit und Geduld der Interviewer hervor und waren mit der flexiblen Terminvereinbarung und -verschiebung sehr zufrieden. Auch zwei der Elternbegleiter hatten von Eltern, die sie zur Teilnahme vermittelt hatten, eine Rückmeldung erhalten. Demnach hätten diese Befragten die Untersuchung interessant gefunden, die Befragungssituation als sehr angenehm und wertschätzend empfunden und seien danach sehr zufrieden und glücklich gewesen. Auch bezüglich ihrer eigenen Befragung äußerten sich diese beiden Elternbegleiter dahingehend positiv, dass die Fragen sinnvoll und umfassend gewesen seien und ihnen sehr gut gefallen hätten. Ein Elternbegleiter beurteilte auch die Methode der Onlinebefragung als gut, da man die Umfrage dann flexibel und in Ruhe durchführen könne, wenn man Zeit dafür habe. Hinsichtlich der Organisation wurden zudem die zeitlichen Abstände der Befragungen sowie die Aufrufe dazu per Mail lobend erwähnt.

Eine einzige negative Rückmeldung ergab sich aus der Stellungnahme eines Kursleiters. Dieser war von einem Elternbegleiter informiert worden, dass ein von jenem zur Befragung vermittelter Elternteil sich sehr kritisch über die Elternbefragung geäußert habe. So seien die Fragen viel zu persönlich gewesen und es sei unterstellt worden, man würde die gefragten Aspekte wie beispielsweise die Erziehung nicht hinbekommen. Während diese Person den Fragebogen somit offenbar als sehr defizitorientiert wahrnahm, gab ein Stellungnehmer der Elterngruppe hingegen die Rückmeldung, dass die Fragen vielmehr sogar Anregungen enthalten hätten. Auch der Kursleiter selbst bemerkte, dass diese negative Kritik ein Einzelfall gewesen sei. Da im Rahmen der Datenerhebung hunderte Eltern befragt wurden und den Interviewern dabei kein derartiges Feedback gegeben wurde (obwohl am Ende jedes Interviews offen nachgefragt wurde, ob der Befragte noch Hinweise oder Anmerkungen jedweder Art habe), sowie aufgrund der sehr positiven Rückmeldungen aus den Stellungnahmen, muss dieser einzelne Fall daher in Relation betrachtet werden.

Ein anderer Aspekt, den die Servicestelle in ihrer Stellungnahme ansprach und welcher bereits im Standard N1 Erwähnung fand, bestand in einer zu geringen Absprache mit dem Evaluationsteam hinsichtlich der zeitlichen Planung der Datenerhebungen. Dadurch fanden die Erhebungen, welche die Servicestelle in den Einrichtungen

Elternbegleitung Plus durchführte, teilweise sehr kurz vor oder nach denen der Evaluatoren statt, was hohe Belastungen für die Teilnehmenden bedeutete. Wie ebenfalls im Standard N1 erläutert, hatte es zwar durchaus eine Abstimmung zwischen Evaluationsteam und Servicestelle gegeben, die jedoch aus logistischen Gründen nicht in einer wesentlichen Entzerrung der Erhebungszeiträume resultierte. Zudem wurde auf die dritte Erhebung des Moduls 2 der Evaluation verzichtet, weil die Einrichtungen *Elternbegleitung Plus* durch zwei weitere Erhebungen besonders belastet waren.

Insgesamt waren in der Evaluation von „Elternchance ist Kinderchance" angemessene Verfahren zur Anwendung gekommen, deren Vor- und Nachteile transparent gemacht und hinsichtlich des Zeitplans mit dem Auftraggeber abgestimmt wurden. Das Evaluationsteam war ausreichend qualifiziert und hatte eine Pilotstudie durchgeführt, um die Durchführbarkeit und Verständlichkeit zu erhöhen. Die von den Standards empfohlenen Alternativpläne und Puffer fehlten jedoch weitgehend. Die Verfahren waren mit vertretbarem Aufwand für die Teilnehmenden durchführbar und die Datenerhebung richtete sich an der Verfügbarkeit der Teilnehmenden aus. Entsprechend positiv wurden sie von den Befragten bewertet. Da bezüglich des Aspekts der unnötigen Belastung der Befragten nur sehr wenige Abstriche gemacht werden müssen, kann der Standard somit als überwiegend eingehalten bewertet werden.

⇨ Standard D1 wurde überwiegend erfüllt.

4.2.2.2.2 Diplomatisches Vorgehen (D2)

Evaluationen sollen so geplant und durchgeführt werden, dass möglichst hohe Akzeptanz der verschiedenen Beteiligten und Betroffenen in Bezug auf Vorgehen und Ergebnisse der Evaluation erreicht werden kann.

Indem sie die Interessen aller Beteiligten angemessen berücksichtigen, erhöhen Evaluierende den DeGEval-Standards zufolge die Akzeptanz, Zustimmung und Kooperation der Stakeholder. Dies ist wichtig, um Einschränkungen der Evaluationsaktivitäten sowie Ergebnisverzerrungen oder -missbrauch durch einen oder mehrere Stakeholder

zu vermeiden. Zu diesem Zweck ist es erforderlich, dass die Evaluatoren diplomatisch vorgehen. Besonders im politischen Bereich kommt es häufiger vor, dass Auftraggebende von Evaluationen strategische Zwecke verfolgen und z. B. Entscheidungen, die bereits feststehen, im Nachhinein legitimieren wollen. In solchen Kontexten ist die Förderung des Dialoges zwischen den Entscheidern und anderen Stakeholdern eine wichtige Aufgabe der Evaluatoren. Wenn die verschiedenen Beteiligten zur Kooperation bereit sind, werden sie sich auch wahrscheinlicher an der Evaluation beteiligen, Informationen zur Verfügung stellen und die Ergebnisse anerkennen und nutzen.

Gemäß dem JCSEE-Handbuch sollen Evaluierende sich möglicher Einflussnahmen durch Beteiligte bewusst sein, um ggf. darauf reagieren zu können. Eine solche Einflussnahme kann z. B. darin bestehen, dass eine Gruppe sich einen Vorteil verschafft oder die Evaluation maßgeblich behindert. Da dies die Glaubwürdigkeit der Evaluation erheblich verringern kann, sollten die Evaluatoren auf solche Einflussversuche sensibel reagieren. So sollten sie sich bei einer möglicherweise brisanten Evaluation bereits frühzeitig mit möglichst vielen Beteiligten hinsichtlich deren Standpunkten und Bedenken austauschen und sie von der Fairness des Evaluationsprozesses überzeugen. In einem Vertrag sollten die Bedingungen, unter denen die Evaluation stattfinden soll, niedergeschrieben werden. Auch der Zugriff der Evaluatoren auf die benötigten Daten sowie ihre Zuständigkeit für Redaktion und Verbreitung der Berichte sollten dort festgehalten werden. Der Vertrag sollte zudem, sofern erforderlich, überarbeitet werden können. Der Auftraggeber und andere wichtige Beteiligte sind regelmäßig über den aktuellen Stand der Evaluation zu informieren, damit die Ergebnisse sie nicht gänzlich überraschen und damit die Evaluierenden ihre Reaktionen auf die Berichte erfahren können.

In der hier bewerteten Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ konnte eine überwiegend angemessene Beteiligung der verschiedenen Stakeholder erreicht werden (vgl. Standard N1). Regelmäßige Informationen über den aktuellen Stand der Evaluation erhielt der Auftraggeber in Form halbjährlicher Sachstandsberichte, des Zwischenberichts etwa zur Mitte der Projektlaufzeit sowie mehrerer Präsentationen im Rahmen von Arbeitstreffen. Auch die beiden Träger wurden explizit einbezogen und zu Workshops eingeladen, bei denen ihnen die aktuellsten Ergebnisse vermittelt

sowie Diskussionen ermöglicht wurden. Zu diesem Zweck nahmen weiterhin mehrere Mitglieder des Evaluationsteams an einem Treffen der Kursleiter eines Trägers teil sowie an den beiden Bundeskongressen Elternbegleitung. Bei letzteren standen sie den vielen vertretenen Interessengruppen zum Austausch zur Verfügung. Diese informierenden Aktivitäten fanden nicht so regelmäßig statt wie die Berichte an den Auftraggeber, was jedoch der Anlage der Evaluation als auftraggeberorientiert entsprach.

Es wurden somit seitens der Evaluatoren viele Bemühungen unternommen, um die Beteiligten ausreichend zu berücksichtigen und bei ihnen eine kooperative Haltung sowie Akzeptanz zu erzeugen. Dies resultierte in einer kooperativen Arbeitsbeziehung mit den verschiedenen Stakeholdern. So wurden die Evaluierenden durch den Auftraggeber unterstützt, indem beispielsweise weitere Mittel für Incentives für die schwer erreichbare Kontrollgruppe bereitgestellt wurden. Die Zusammenarbeit mit den Qualifizierungsträgern, welche dem Evaluationsteam die Kontaktierung der zu befragenden Elternbegleiter ermöglichten, war ebenfalls unterstützend und kooperativ, und auch die Fachkräfte, die Eltern für die Befragung vermittelten, zeigten überwiegend eine sehr kooperative Haltung. Hinsichtlich der Akzeptanz der Ergebnisse und deren Nutzung sind nur begrenzte Aussagen möglich, da der umfassende Abschlussbericht zum Zeitpunkt der Stellungnahmen noch nicht freigegeben und nur wenigen Stellungnehmern die Kurzbroschüre mit zentralen Ergebnissen bekannt war. Diejenigen, die die Kurzbroschüre kannten, äußerten sich positiv darüber, was für eine Akzeptanz der Ergebnisse spricht (vgl. Standards N5 und N6). Genutzt wurden diese bislang in erster Linie vom Auftraggeber, indem sie den Überarbeitungsprozess der Qualifizierungshandbücher unterstützten.

Die Evaluatoren waren sich von Beginn an im Klaren darüber, dass die Beteiligten am Bundesprogramm teilweise gegenläufige Interessen verfolgten und daher Einflussnahmen möglich waren. So wurden die Qualifizierungen zu Elternbegleitern von zwei Trägern durchgeführt, die in einem konkurrierenden Verhältnis zueinander standen. Beide schienen ein legitimatorisches Interesse an der Evaluation zu haben, und zwar dahingehend, dass eine möglichst positive Bewertung des eigenen Curriculums erhofft wurde. Die unterschiedlichen Haltungen zeigten sich z. B. darin, dass ein Träger sich eine vergleichende Auswertung der beiden Curricula erhofft hatte, der andere hingegen nicht

(vgl. Standard N5). Die Evaluatoren bezogen daher beide Träger gleichermaßen ein und ließen ihnen die gleichen Informationen zukommen. Wie erwähnt, resultierte dies in einer guten Kooperation und es wurden keine direkten Einflussversuche wahrgenommen. Indirekte Beeinflussungen fanden dahingehend statt, dass die Träger die Evaluatoren von den Vorteilen ihres jeweiligen Curriculums zu überzeugen versuchten, was allerdings auch naheliegend und verständlich erscheint. Ein Träger legte zudem eine etwas skeptische Haltung insbesondere den quantitativen Daten der Evaluation gegenüber an den Tag, möglicherweise mit der Intention, dadurch einer potentiellen Kritik vorzubeugen. Obwohl die Evaluatoren sich bemühten, diese Bedenken auszuräumen bzw. die Vor- und Nachteile der Datenerhebungsmethoden zu erläutern, schienen sich diese Zweifel nicht gänzlich auflösen zu lassen. Dass es solche Vorbehalte gab, war eine wichtige Information für die Evaluierenden. Sie wurden im Rahmen der Arbeitstreffen bekannt, was den Vorteil dieses offenen Austausches verdeutlicht.

Das Evaluationsteam hatte somit den Austausch mit verschiedenen Beteiligten verfolgt und dabei deren Standpunkte erfragt sowie sie über den aktuellen Stand der Evaluation informiert. Mögliche Bedenken der Stakeholder hätten jedoch durchaus direkter angesprochen werden können, um auch eventuell vorhandene Befürchtungen in Erfahrung zu bringen. In diesem Zusammenhang hätte auch der Dialog zwischen den Auftraggebenden und weiteren Beteiligten gefördert werden können. Wie beim Standard N2 erläutert, sollte im Rahmen der Evaluation ursprünglich ein Vergleich der Curricula vorgenommen werden, jedoch entschied sich der Auftraggeber dann zunächst gegen solche vergleichenden Auswertungen. Da diese sich im Evaluationsverlauf jedoch als sinnvoll erwiesen und vom Evaluationsteam empfohlen wurden, wurden sie letztendlich doch vom Auftraggeber gewünscht und von den Evaluatoren durchgeführt. Die Evaluatoren hätten einen diesbezüglichen Austausch zwischen Auftraggebenden und Trägern empfehlen können, damit die Richtung der Evaluation sowie eine Anpassung derselben klar kommuniziert würde. Dies hätte verhindert, dass ein Träger die Erwartung hatte, dass Unterschiede der Curricula geprüft und berichtet würden, und dass diese Erwartung enttäuscht wurde (vgl. Standard N1).

Seitens des Auftraggebers wurde Einflussnahme dahingehend wahrgenommen, dass dieser sämtliche Kommunikation nach außen, also

Berichte und Präsentationen, prüfte, was auch vertraglich gestützt war. Hinsichtlich der Präsentationen ergaben sich wenige Änderungswünsche, während bei der Kurzbroschüre und beim Abschlussbericht relativ viele Korrekturen erbeten wurden. Diese wurden vom Evaluationsteam weitgehend für sinnvoll und nachvollziehbar befunden, es gab jedoch auch kritische Punkte, wenn z. B. Informationen, die die Evaluatoren als zentral erachteten, gestrichen werden sollten oder das Wording positiver gewünscht wurde, als die Ergebnisse es rechtfertigten (vgl. hierzu auch Standard F4). Die Evaluierenden verfolgten hier eine diplomatische Strategie, indem sie die Korrekturen des Auftraggebers akzeptierten, sofern sich dies mit den Ergebnissen und den wissenschaftlichen Standards der Evaluatoren vereinbaren ließ, sie jedoch auch deutlich machten, welche Änderungen nicht umgesetzt werden konnten und dies begründeten.

Von den vertraglichen Vereinbarungen, die im Handbuch empfohlen werden, waren die meisten in der vorliegenden Evaluation vorhanden. Die Evaluationsbedingungen waren ebenso wie der Zugriff auf die Daten, die redaktionelle Verantwortlichkeit und die Möglichkeit der Vertragsanpassung vertraglich festgehalten. In Bezug auf die Verbreitung der Berichte waren die Rechte vereinbart worden, nicht jedoch die Pflicht der Ergebnisverbreitung (vgl. Standards N7 und N8). Dem Auftraggeber zufolge sei dies allerdings als Merkmal der Auftragsforschung zu sehen. So sei der Auftraggeber zentraler Nutzer der Evaluationsergebnisse und dieser entscheide über die weitere Verbreitung.

Insgesamt lässt sich feststellen, dass einige, aber nicht alle Anforderungen dieses Standards erfüllt wurden. Die verschiedenen Beteiligten waren überwiegend angemessen in die Evaluation einbezogen sowie über den aktuellen Stand auf dem Laufenden gehalten worden. Den Evaluatoren war es gelungen, eine kooperative Haltung zu erzeugen und sie waren sich möglicher Einflussnahmen bewusst und ergriffen entsprechende diplomatische Maßnahmen. Beim Austausch mit den Stakeholdern hätten deren Bedenken stärker thematisiert, und ebenso hätte der Dialog zwischen Auftraggeber und Trägern mehr gefördert werden können. Die vertraglichen Vereinbarungen entsprachen überwiegend den Empfehlungen, lediglich die Verbreitung der Berichte war nicht vollständig abgedeckt. Zusammengenommen kann der Standard

als teilweise bis überwiegend eingehalten bewertet werden, wobei die konservativere Einschätzung den Vorzug erhält.

⇨ Standard D2 wurde teilweise erfüllt.

4.2.2.2.3 Effizienz von Evaluation (D3)

Der Aufwand für Evaluation soll in einem angemessenen Verhältnis zum Nutzen der Evaluation stehen.

Die Bestimmung von Kosten und Nutzen einer Evaluation stellt ein schwieriges Vorhaben dar, wie die DeGEval-Standards erläutern. Die Kosten beinhalten den finanziellen Wert aller aufgewendeten Ressourcen. Darunter fallen beispielsweise der zeitliche Aufwand aller Stakeholder, die Bezüge der Evaluatoren sowie Reise- und Materialkosten. Auch von Dritten übernommene Kosten müssen hier mit einbezogen werden und es gibt Kosten, die sich als nicht quantifizierbar erweisen. Der Nutzen umfasst die direkten und indirekten Effekte der evaluierten Maßnahme und kann neben intendierten Auswirkungen auch unbeabsichtigte Nebenwirkungen beinhalten. Die Ermittlung des Nutzens gestaltet sich noch herausfordernder als die der Kosten und kann üblicherweise nur anhand von Schätzungen erfolgen. Trotz dieser Schwierigkeiten sollten Kosten und Nutzen einer Evaluation stets abgewogen werden, vor allem, wenn zu entscheiden ist, ob eine Evaluation durchgeführt wird oder nicht. Daher sollte der Evaluationsplan eine nachvollziehbare Übersicht über die zu erwartenden Kosten und Nutzen enthalten.

Im JCSEE-Handbuch wird weiterhin erwähnt, dass eine Evaluation dann effizient ist, wenn ihr Nutzen gleich den Kosten ist oder diese übersteigt. Wenn mehrere Evaluationspläne vorliegen, sollte derjenige ausgewählt werden, der das günstigste Verhältnis von Kosten und Nutzen aufweist. Wenn das aufwendigste Evaluationsdesign die besten Informationen liefert, sollten die – verglichen mit den anderen Designs – zusätzlichen Daten die höheren Kosten aufwiegen. Ist der Nutzen mehrerer Evaluationspläne gleich hoch einzuschätzen, sind allein die Kosten ausschlaggebend. Auch im Handbuch wird betont, wie schwierig die Festlegung insbesondere des Nutzens einer Evaluation ist. Doch selbst wenn diese Parameter nicht vollständig ermittelt werden können, sollte dennoch darauf geachtet werden, dass Ressourcen

sorgfältig eingesetzt werden und auf ein effizientes Vorgehen geachtet wird.

In der vorliegenden Evaluation wurde ein umfangreiches und eher aufwendiges Design umgesetzt, welches detaillierte Daten lieferte, um die Fragestellungen zu beantworten. Im Projektangebot wurden die dafür eingeplanten Ressourcen aufgeschlüsselt und somit die erwarteten Kosten kalkuliert. Da die Angebote der weiteren Bieter nicht bekannt sind, kann nicht beantwortet werden, ob die vorliegende Evaluation den Ansatz mit den niedrigsten Kosten darstellte oder ob das Design und die damit erreichbaren Informationen die möglicherweise höheren Kosten aufgewogen hatten. Für letzteres könnte eine Äußerung im Rahmen der Stellungnahme des Auftraggebers sprechen, wonach das plausible Design und die Erhebungsinstrumente eine wichtige Rolle bei der Vergabe der Evaluation an DJI und FAU gespielt hätten.

Die Ressourcen wurden sorgfältig eingesetzt und damit ein effizientes Vorgehen verfolgt. Es wurde auf eine günstige Durchführung der Testungen geachtet, die dennoch zu den situativen Anforderungen passte. So wurde bei den Fachkräften die Methode der Online-Befragung gewählt, während die Eltern telefonisch befragt wurden, was zwar für das Evaluationsteam deutlich aufwendiger war, aber einen höheren Rücklauf versprach. Die qualitativen Befragungen der Eltern wurden ebenfalls als telefonische Interviews durchgeführt, weil der zusätzliche Informationsgewinn einer face-to-face-Situation nicht ausreichend mehr Informationen erbracht hätte, um die dadurch entstehenden Kosten aufzuwiegen. Für die Fokusgruppen im Modul 2 war hingegen die Durchführung vor Ort zentral, weshalb hier die zeitlichen und finanziellen Aufwendungen für die Anreise der Evaluatoren in Kauf genommen wurden. Bei den qualitativen Interviews mit Fachkräften im Rahmen des Moduls 1 wurden einige face-to-face-Interviews realisiert, die übrigen Gespräche wurden telefonisch geführt, um Kosten und Zeit zu sparen. Es wurden somit in allen drei Modulen Kosten und Nutzen der jeweiligen Methoden gewissenhaft abgewogen.

Ein zentraler Nutzen der Evaluation bestand in Empfehlungen zur Überarbeitung der Weiterbildungscurricula für die zweite Phase des Bundesprogramms, um diese optimieren und die Ressourcen in der nächsten Phase noch zielgenauer einsetzen zu können. Darüber hinaus

wurden viele für Wissenschaft, Politik und Fachpraxis relevante Erkenntnisse zu diesem neuen Programm gewonnen, beispielsweise wie Elternbegleitung in der Praxis tatsächlich umgesetzt wird und wie die Eltern sie wahrnehmen. Wenn diese Ergebnisse, z. B. in Form des Abschlussberichts, noch weiter verbreitet werden, könnten sich zudem weitere Nutzungen ergeben, die dann auch den Nutzen der vorliegenden Evaluation erhöhen würden.

Die Frage, ob der Nutzen die Kosten ausgleicht bzw. übersteigt und die Evaluation damit als effizient einzustufen ist, kann nicht definitiv beantwortet werden. Auch der Auftraggeber äußerte in seiner Stellungnahme Zweifel daran, ob man den Nutzen quantifizieren könne und solle. Er bezog sich stattdessen auf den großen Anklang, den die Evaluation bei vielen Wissenschaftlern gefunden hatte (bei der wissenschaftlichen Abschlusstagung in Berlin am 5.12.2014, vgl. Standard F3) und bemerkte dazu, dies spreche für sich.

Da insgesamt einige Aspekte für eine Kostenwirksamkeit der Evaluation von „Elternchance ist Kinderchance“ sprechen und die Evaluatoren einen sorgfältigen Umgang mit den Ressourcen an den Tag gelegt hatten, wird dieser Standard als erfüllt angesehen.

⇨ Standard D3 wurde erfüllt.

4.2.2.3 Fairnessstandards

Die Fairnessstandards sollen sicherstellen, dass in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird.

4.2.2.3.1 Formale Vereinbarungen (F1)

Die Pflichten der Vertragsparteien einer Evaluation (was, wie, von wem, wann getan werden soll) sollen schriftlich festgehalten werden, damit die Parteien verpflichtet sind, alle Bedingungen dieser Vereinbarung zu erfüllen oder aber diese neu auszuhandeln.

Die DeGEval-Standards raten dazu, bei einer Evaluationsstudie eine formale Vereinbarung zu treffen, die mindestens die Aspekte der Finanzen, Zeit, Methodik und mitwirkenden Personen beinhaltet und insbesondere die Rechte und Pflichten der Beteiligten festlegt. Gegenseitiger Respekt und Vertrauen der Parteien sollten dabei die Grundlage darstellen. Eine in Schriftform festgehaltene Absprache kann Missverständnissen vorbeugen oder zumindest zu deren rascherer Klärung beitragen. Es gilt zu beachten, dass zu strikte Vereinbarungen, z. B. in Bezug auf Detailfragestellungen oder die Methodik, den Evaluationsprozess und damit den Erkenntnisgewinn behindern können. Im Evaluationsverlauf können Änderungen notwendig sein und eine Neuverhandlung der Vertragsbedingungen erforderlich machen. Dies wird erleichtert, indem die Evaluierenden die Vertragspartner von Beginn an auf die Möglichkeit solcher Anpassungen hinweisen.

Das Handbuch zu den JCSEE-Standards differenziert die möglichen Vertragsinhalte. Demnach sollten in der Vereinbarung folgende Punkte angesprochen werden:

a) Verantwortlichkeiten des Personals der zu evaluierenden Einrichtung bei der Durchführung der Evaluation sowie bei der Teilnahme an der Evaluation
b) Zwecke der Evaluation
c) Ermittlung der Beteiligten und Betroffenen sowie der zu untersuchenden Fragestellungen
d) Genaue Beschreibung der zu erbringenden Leistungen
e) Verfahren der Datenerhebung, einschließlich Informationsquellen, Auswahl der Teilnehmenden, Instrumente und anderen Techniken zur Datenerhebung
f) Verfahren zur Analyse von quantitativen und qualitativen Daten und Informationen
g) Managementplan (z. B. Pflichtenheft mit Zuweisung der Verantwortlichkeiten)
h) Plan für die Berichterstattung einschließlich Überlegungen zu Format und Abgabe von Zwischenberichten und des Schlussberichts (Anzahl und Art der Berichte, Länge, Adressaten, Darstellungsmethoden); Anonymität der Befragten; Überprüfung von Berichtsentwürfen; Gegendarstellung durch diejenigen, die evaluiert wurden; Zuständigkeit für die Redaktion und endgültige Freigabe der fertiggestellten Berichte

i) Methoden der Qualitätskontrolle bei der Datenerhebung, Auswertung und Berichterstattung
j) Dienstleistungen, Personal, Informationen und Materialien, die vom Auftraggeber bereitgestellt werden – inklusive Datenzugang
k) Zeitplan für die Arbeit von Auftraggebern und Evaluatoren
l) Verfahren zur Änderung und Beendigung des Vertrags
m) Budget für die auszuführende Arbeit, einschließlich Beträgen, die unmittelbar nach der Erledigung bestimmter Aufgaben oder zu bestimmten Zeitpunkten zu zahlen sind
n) Festlegung von regelmäßigen Überprüfungen und Modifikationen der Vereinbarung, wenn im Verlauf der Arbeit Veränderungen wünschenswert erscheinen

Weiterhin erwähnt das Handbuch, dass Änderungen der Vereinbarung zu verhandeln und dokumentieren sind, wenn sich aufgrund geänderter Umstände Abweichungen beim Arbeitsumfang, den Kosten oder dem Zeitplan ergeben. Zudem kann es sinnvoll sein, den Vertrag von Dritten auf Klarheit und Angemessenheit hin prüfen zu lassen.

Die Evaluation des Bundesprogramms „Elternchance ist Kinderchance" wurde von DJI und FAU gemeinsam durchgeführt. Beide Institutionen schlossen jeweils einen Werkvertrag mit dem Auftraggeber, dem BMFSFJ ab. Dieser umfasste eine Vielzahl an Vereinbarungen und nahm Bezug auf das Angebot der Auftragnehmer sowie auf die Kostenkalkulation. Zwischen DJI und FAU wurde eine Kooperationsvereinbarung getroffen und zudem eine Vereinbarung zum Datenaustausch unterzeichnet. Die genauen Inhalte dieser Übereinkünfte im Detail darzustellen, ist aus Platzgründen nicht möglich. Sie lassen sich dahingehend zusammenfassen, dass die meisten der oben genannten Punkte in den Verträgen enthalten waren. Es werden nachfolgend diejenigen Aspekte benannt, welche *nicht* Bestandteil der vertraglichen Vereinbarungen waren.

a) Verantwortlichkeiten des Personals der zu evaluierenden Einrichtung bei der Durchführung der Evaluation sowie bei der Teilnahme an der Evaluation: Da es sich um Vereinbarungen zwischen Evaluationsteam und Auftraggeber bzw. zwischen

den beiden Evaluatoren handelte, waren keine Bestimmungen zu den Teilnehmenden an der Evaluation enthalten.

c) Ermittlung der Beteiligten und Betroffenen sowie der zu untersuchenden Fragestellungen: Während die Fragestellungen im Angebot der Evaluatoren angeführt waren, fehlte eine Übereinkunft zur Ermittlung der Beteiligten und Betroffenen.

h) Plan für die Berichterstattung einschließlich Überlegungen zu Format und Abgabe von Zwischenberichten und des Schlussberichts (Anzahl und Art der Berichte, Länge, Adressaten, Darstellungsmethoden); Anonymität der Befragten; Überprüfung von Berichtsentwürfen; Gegendarstellung durch diejenigen, die evaluiert wurden; Zuständigkeit für die Redaktion und endgültige Freigabe der fertiggestellten Berichte: Die meisten der genannten Gesichtspunkte waren vereinbart worden, es fehlte jedoch eine Festlegung der Adressaten der Berichte, der Darstellungsmethoden sowie der Gegendarstellung durch die Evaluierten.

i) Methoden der Qualitätskontrolle bei der Datenerhebung, Auswertung und Berichterstattung: Während bei der Evaluation sehr auf Qualität geachtet und diese durch viele Maßnahmen sichergestellt wurde (z. B. durch Schulungen für die Bereiche der Datenerhebung und -auswertung), war diese nicht vertraglich geregelt.

j) Dienstleistungen, Personal, Informationen und Materialien, die vom Auftraggeber bereitgestellt werden – inklusive Datenzugang: Da die Evaluatoren die Daten bei anderen Beteiligten erhoben und diesbezüglich nicht auf eine Bereitstellung durch den Auftraggeber angewiesen waren, fehlte ein entsprechender Passus im Vertrag. Wenn in diesem Zusammenhang Kooperation notwendig war, klappte diese jedoch reibungslos, indem z. B. der Auftraggeber die Datenerhebung mit einem Schreiben unterstützte, bei dem um Teilnahme geworben wurde.

Insgesamt waren somit von den 14 im Handbuch des JCSEE aufgezählten Vertragsinhalten neun Aspekte auch im Werkvertrag der Evaluation von „Elternchance ist Kinderchance“ berücksichtigt worden, weitere zwei immerhin teilweise und drei fehlten. Unter der Voraussetzung, dass nicht alle Punkte bei jeder Evaluation passend sind (was

für vorliegende Evaluation im Diskussionsteil erörtert wird), kann die vertragliche Vereinbarung damit als den Standards sehr gut entsprechend eingestuft werden. Sie enthielt zudem den Hinweis, dass die Vertragserfüllung in enger Abstimmung und vertrauensvoller Kooperation der beiden Parteien geschehen soll.

Schließlich ist zu erwähnen, dass eine Prüfung des Vertrags durch Dritte erfolgte, so z. B. im Falle der FAU, welche den Vertrag durch den Justiziar der Universität prüfen und daraufhin noch einen Passus ergänzen ließ, der die Möglichkeit der Erstellung von Qualifizierungsarbeiten im Rahmen des Projekts zusicherte.

Es lässt sich festhalten, dass in den Verträgen der vorliegenden Evaluation sehr viele der von den Standards empfohlenen Punkte enthalten waren und die Vereinbarung auf gegenseitigem Vertrauen aufbaute. Es fehlten nur wenige Aspekte, darunter die Verbreitung des Berichts, die zu vereinbaren wichtig gewesen wäre, was in der vorliegenden Meta-Evaluation auch bei anderen Standards Erwähnung findet. Der Standard wird daher als überwiegend erfüllt betrachtet.

⇨ Standard F1 wurde überwiegend erfüllt.

4.2.2.3.2 Schutz individueller Rechte (F2)

> Evaluationen sollen so geplant und durchgeführt werden, dass Sicherheit, Würde und Rechte der in eine Evaluation einbezogenen Personen geschützt sind.

Den DeGEval-Standards zufolge ist es wichtig, dass Evaluierende die Würde und Selbstachtung der Menschen, mit denen sie im Rahmen der Evaluation zu tun haben, nicht verletzen. Gerade wenn Evaluationen auch Schwächen herausfinden, kann der Interessenschutz beeinträchtigt werden, sodass dann die Interessen besonders sorgfältig abgewogen werden müssen. Dies gilt in besonderem Maße, wenn Verstöße gegen Recht und Gesetz offengelegt werden.

Die Erläuterungen der deutschen Standards sind recht knapp gehalten. Dabei umfasst der Standard F2 sogar zwei der amerikanischen Standards („Schutz individueller Menschenrechte“ und „Human gestaltete Interaktion“). Das JCSEE-Handbuch benennt als zu beachtende Rechte neben gesetzlich verankerten Bestimmungen auch solche

Rechte, die sich auf ethische Prinzipien, den gesunden Menschenverstand oder Regeln der Höflichkeit gründen. Dazu zählen unter anderem die Freiwilligkeit der Teilnahme, das Recht auf Rücktritt ohne negative Konsequenzen und die vertrauliche Behandlung von Daten. Wenn Evaluatoren die Persönlichkeitsrechte nicht beachten und dadurch Teilnehmende schädigen, kann dies die Nutzung der Evaluationsergebnisse beeinträchtigen und auch gerichtliche Folgen nach sich ziehen. Um die Anforderungen dieses Standards zu erfüllen, sollten Evaluatoren die jeweiligen Gesetze und ethischen Richtlinien kennen. Sie sollten die Werte der Beteiligten ermitteln (vgl. Standard N5) und auf deren mögliche Bedenken eingehen (vgl. Standard D2). Im Falle von Sprachbarrieren sollten sie für eine angemessene Verständigung mit den Teilnehmenden sorgen. Bei der Datenerhebung sollte die Störung der Befragten möglichst gering gehalten werden (vgl. Standard D1). Die Beachtung der Rechte der Teilnehmenden sollte in einer schriftlichen Übereinkunft festgehalten werden und die Teilnehmenden selbst sind über ihre Rechte zu informieren. Zudem sollten die Evaluationsberichte keinen Rückschluss auf die Identität der Befragten ermöglichen.

In der vorliegenden Evaluation wurden Sicherheit, Würde und Rechte der beteiligten Personen durchweg gewährleistet. Die Studie war von vornherein in diesem Sinne angelegt und datenschutzrechtliche Bestimmungen sowie die Beachtung des Gender Mainstreaming waren Bestandteil des Werkvertrages. Das Evaluationsteam war über Gesetze und ethische Richtlinien informiert. Die Freiwilligkeit der Teilnahme und die Vertraulichkeit im Umgang mit den erhobenen Daten hatten für die Evaluatoren oberste Priorität. Dies lässt sich durch folgendes Beispiel veranschaulichen: Der Zugang zu den Fachkräften sollte ursprünglich anhand der Bewerbungen der Teilnehmer erfolgen, die der Servicestelle vorlagen und über die die Kontaktdaten der Bewerber an die Evaluatoren weitergeleitet werden sollten. Da in den Bewerbungen jedoch keine Einverständniserklärung zur Weitergabe von Daten enthalten war, wurde dieses Vorgehen als nicht datenschutzkonform betrachtet. Der Zugang erfolgte daher über die Qualifizierungsträger, wobei für die beiden Träger ein jeweils unterschiedliches Vorgehen gewählt werden musste, was in einer Verdoppelung der Anzahl der Befragungen im Modul 1 und in einem deutlich erhöhten Zeitaufwand für die Evaluatoren resultierte (vgl. Standard N7). Es mussten somit

zugunsten der Fairness Abstriche bei der Nützlichkeit in Kauf genommen werden.

Weiterhin kommunizierten die Evaluierenden den Teilnehmenden ihre Rechte offen. Dies geschah z. B. mittels Flyern und Datenschutzerklärungen, auf denen über die Teilnahmebedingungen, die Freiwilligkeit der Teilnahme, die Vertraulichkeit der Daten etc. informiert wurde. Zur Umgehung von Sprachbarrieren wurden im Rahmen der Elternbefragung Interviews auch in anderen Sprachen als Deutsch ermöglicht. Wie bereits bei den Standards N5, D1 und D2 erläutert, waren die Werte der Beteiligten offenbar angemessen berücksichtigt, ihre Bedenken ebenfalls zumindest teilweise ermittelt worden und die Datenerhebung war ganz überwiegend ohne unnötige Störung der Befragten erfolgt. Auch die Evaluationsberichte ließen keine Identifizierung der Teilnehmenden zu.

Die Stellungnahmen der Vertreter von Elternbegleitern und Eltern, also den Gruppen der Befragten, machten deutlich, dass diese sämtlich sehr zufrieden mit den Befragungen und dem Vorgehen des Evaluationsteams dabei gewesen waren. Die von einem Kursleiter berichtete Unzufriedenheit einer Teilnehmerin aus der Elterngruppe ist in Relation zu den vielen positiven Rückmeldungen zu betrachten und bezog sich zudem auf die Inhalte der Befragung, nicht auf verletzte Persönlichkeitsrechte (vgl. Standard D1).

Die intensiven und erfolgreichen Bemühungen der Evaluierenden, die Würde und Rechte der Beteiligten zu schützen, ergeben eine positive Bewertung des Standards F2.

⇨ Standard F2 wurde erfüllt.

4.2.2.3.3 Vollständige und faire Überprüfung (F3)

Evaluationen sollen die Stärken und die Schwächen des Evaluationsgegenstandes möglichst vollständig und fair überprüfen und darstellen, sodass die Stärken weiter ausgebaut und die Schwachpunkte behandelt werden können und die Rechte der in eine Evaluation einbezogenen Personen geschützt sind.

In Evaluationen sollen, so fordern es die DeGEval-Standards, nicht nur die Schwächen des Evaluandums, sondern auch seine Stärken

untersucht und beschrieben werden. Beides wird für eine vollständige und ausgewogene Bewertung des Evaluationsgegenstandes als wichtig erachtet. Denn es kann sein, dass Maßnahmen zur Behebung von Schwächen eine Beeinträchtigung der Stärken mit sich bringen. Demgegenüber können vorhandene Stärken aber auch als Ausgleich für gefundene Schwächen dienen. Oft haben Evaluationen neben der Verringerung von Schwächen auch den Ausbau von Stärken zum Ziel. Zudem wirkt sich die Erfassung der Stärken positiv auf die Kooperation der Beteiligten und ihre Akzeptanz der Evaluation aus. Stärken und Schwächen sollten so vollständig wie möglich überprüft und dargestellt werden, wobei alle wichtigen Aspekte behandelt werden sollen und auf Fairness gegenüber den Beteiligten zu achten ist. Es sollten durch den Auftraggeber oder weitere Stakeholder keine Einflussversuche auf die Evaluation oder ihren Bericht unternommen werden. Die Evaluierenden sind für eine ausgewogene, faire und vollständige Evaluation verantwortlich.

Im Handbuch des JCSEE wird ergänzend darauf eingegangen, dass auch die Methoden der Evaluation einschließlich ihrer Vor- und Nachteile dargestellt werden sollten, damit deren Bedeutung für die Ergebnisse und Schlussfolgerungen nachvollziehbar wird. Wenn Daten nicht zugänglich sind, die für die Evaluation relevant wären, sollte dies thematisiert und die resultierenden Auswirkungen dieser fehlenden Daten auf die Bewertung sollten abgeschätzt werden. Zudem sollte von sachverständigen Stellen mit unterschiedlichen Perspektiven kritische Rückmeldung bezüglich der Vollständigkeit und Fairness der Bewertung eingeholt werden.

Bei der hier beschriebenen Evaluation wurden im Rahmen des Abschlussberichts sowohl Schwächen und Verbesserungsbedarfe als auch Stärken des Programms „Elternchance ist Kinderchance“ berichtet. Dabei wurde auf Vollständigkeit und Berücksichtigung aller Perspektiven geachtet, um eine faire Darstellung zu gewährleisten. Auch die Datenerhebung war dementsprechend angelegt worden. So waren beispielsweise in die qualitativen Interviews im Modul 1 gezielt sowohl Personen einbezogen worden, die mit der Qualifizierung zufrieden gewesen waren, als auch solche, die sich weniger zufrieden geäußert hatten, um einerseits Stärken hervorheben und andererseits auch Verbesserungsmöglichkeiten aufzeigen zu können.

Der Abschlussbericht enthält weiterhin eine differenzierte Schilderung der verwendeten Methoden sowie einige Gedanken zur Methodenkritik, wobei letztere vergleichsweise knapp gehalten sind. Da in diesem Bericht die Daten der Kontrollgruppe der Elternstichprobe fehlten und später in einem separaten Bericht nachgereicht wurden, war ein entsprechender Hinweis auf die Vorläufigkeit der Ergebnisse enthalten. Kritische Kommentare durch Sachverständige erhielt das Evaluationsteam einerseits beim Abschluss-Workshop, bei welchem Experten unterschiedlicher Disziplinen ihre Eindrücke von der Evaluation und ihren Ergebnissen äußerten. Diese waren vielfach lobend sowie durchweg konstruktiver Natur und einige der erhaltenen Anregungen konnten im Abschlussbericht noch umgesetzt werden. Andererseits wurden sowohl die Kurzbroschüre als auch der Abschlussbericht von einer sachverständigen und unabhängigen Person im DJI gelesen und freigegeben. Einflussnahmen waren im Evaluationsverlauf kaum zu verzeichnen, lediglich der starke Mitsprachebedarf der Auftraggebenden bei zur Veröffentlichung bestimmten Berichten ist zu erwähnen. Wie im Standard D2 erläutert, verfuhren die Evaluatoren diesbezüglich diplomatisch und verteidigten, falls notwendig, auch ihre eigene Position sowie die Interessen anderer Beteiligter. Den ebenfalls im Standard D2 erwähnten Beeinflussungen durch die Qualifizierungsträger, welche das Evaluationsteam von den Vorteilen ihres jeweiligen Curriculums zu überzeugen versuchten, wurde ebenfalls diplomatisch begegnet, indem die Evaluatoren stets beide Träger gleichermaßen einbezogen und in den Erhebungsinstrumenten beide Curricula in gleichem Maße Berücksichtigung fanden.

Für eine abschließende Bewertung dieses Standards wäre eine Beurteilung der Vollständigkeit und Fairness des Abschlussberichts durch verschiedene Stakeholder wichtig. Im Rahmen der Stellungnahmen konnte diese jedoch nicht vorgenommen werden und auch zur Kurzbroschüre konnten nur einige Befragte Aussagen machen (vgl. Standard N6). Diese äußerten sich positiv über deren Verständlichkeit sowie Fairness der Darstellung.

Insgesamt wurden also nach Einschätzung des Evaluationsteams im Abschlussbericht in vollständiger und fairer Weise Stärken und Schwächen des Bundesprogramms dargelegt. Ob andere Beteiligte diese Bewertung teilen, kann an dieser Stelle noch nicht beantwortet werden, sodass auch die Beurteilung des Standards als vorläufig anzusehen

ist. Die übrigen hier skizzierten Informationen sprechen für eine Einhaltung der Anforderungen des Standards F3, wenngleich die Methodenkritik im Abschlussbericht als relativ knapp einzustufen ist. Aus diesem Grund ergibt sich folgende (vorläufige) Bewertung:

⇨ Standard F3 wurde überwiegend erfüllt.

4.2.2.3.4 Unparteiische Durchführung und Berichterstattung (F4)

> Die Evaluation soll unterschiedliche Sichtweisen von Beteiligten und Betroffenen auf Gegenstand und Ergebnisse der Evaluation in Rechnung stellen. Berichte sollen ebenso wie der gesamte Evaluationsprozess die unparteiische Position des Evaluationsteams erkennen lassen. Bewertungen sollen fair und möglichst frei von persönlichen Gefühlen getroffen werden.

Zu diesem Standard wird in den Richtlinien der DeGEval ausgeführt, dass bei Evaluationen viele, oftmals unterschiedliche Sichtweisen der verschiedenen Stakeholder vorkommen und die Möglichkeit in Betracht zu ziehen ist, dass die Evaluation von einer Gruppe vereinnahmt oder instrumentalisiert wird. Die Evaluation sollte nicht einer einzigen Perspektive folgen, sondern versuchen, alle Interessen fair zu behandeln. Aus diesem Grund sollten die Evaluatoren eine möglichst unabhängige Stellung einnehmen und weder zu den Auftraggebenden noch zu anderen Stakeholdern ein zu enges Verhältnis pflegen. Die Beziehung, in der die Evaluierenden zum Auftraggeber sowie zu weiteren relevanten Beteiligten stehen, sollte zu Beginn geklärt werden. Dabei sollte auch die Veröffentlichung von Evaluationsberichten festgelegt werden.

Auch dieser Standard umfasst zwei JCSEE-Richtlinien („Deklaration von Interessenkonflikten“ und „Unparteiische Berichterstattung“). In diesen wird zunächst der Begriff des Interessenkonflikts näher erläutert, welcher vorliegt, wenn persönliche oder finanzielle Interessen der Evaluatoren entweder die Evaluation beeinflussen oder von ihr beeinflusst werden. Dies kann der Fall sein, wenn die Evaluierenden einen Nutzen aus den Evaluationsergebnissen ziehen können, indem z. B. bei positiven Ergebnissen ihre eigenen Verträge verlängert werden bzw. sie dann wahrscheinlicher auch weitere Aufträge bekommen werden. Doch auch bei anderen Beteiligten können Interessenkonflikte

vorliegen, wenn die Evaluation oder die evaluierte Maßnahme von unverhältnismäßig großer persönlicher oder finanzieller Bedeutung für sie sind. In solchen Fällen können Beeinflussungsversuche auf den Evaluationsprozess oder auf die Ergebnisse durch diese Beteiligten vorkommen. Es gibt somit viele denkbare Varianten von Interessenkonflikten, welche sowohl bei internen als auch bei externen Evaluationen auftreten können. Entscheidend ist den amerikanischen Richtlinien zufolge, wie mit vorhandenen Interessenkonflikten umgegangen wird. Dabei sollten mögliche Ursachen bereits in den ersten Besprechungen mit dem Auftraggeber ermittelt und klar beschrieben sowie vorbeugende Maßnahmen schriftlich festgehalten werden. Personen, die eine andere Sicht auf die Evaluation haben, sollten einbezogen und so die Offenheit für andere Auffassungen bewahrt werden. Die Vor- und Nachteile, die sich für verschiedene Beteiligte aus der Evaluation ergeben könnten, sind abzuschätzen und Maßnahmen gegen eine mögliche Beeinflussung durch diese Gruppen zu treffen. Des Weiteren ist auf unparteiische Berichterstattung zu achten. So sollen die Berichte alle relevanten Sichtweisen berücksichtigen und nicht eine mögliche Interpretation anderen vorziehen.

Auch bei der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ waren verschiedene und teilweise divergierende Sichtweisen vorhanden, so z. B. die der beiden konkurrierenden Qualifizierungsträger, von deren Seite auch indirekte Beeinflussungsversuche wahrgenommen wurden (vgl. Standard D2). Die Sichtweisen der Stakeholder sowie mögliche Vor- und Nachteile für die beteiligten Gruppen wurden frühzeitig ermittelt und Maßnahmen erdacht, um möglichen Beeinflussungen zu begegnen. So wurde der Austausch mit beiden Trägern gleichermaßen gepflegt, um keine einseitigen Einflussnahmen zu ermöglichen und umgekehrt den Trägern zu vermitteln, dass keine der beiden Sichtweisen bevorzugt wurde. Diese vorbeugenden Maßnahmen wurden allerdings nicht, wie von den Standards empfohlen, schriftlich festgehalten, was sich auch nicht als notwendig erwies. Mit dem Auftraggeber bestand eine enge Zusammenarbeit, was damit zu begründen ist, dass es sich um eine auftraggeberorientierte Evaluation handelte. Von dieser Seite war ebenfalls eine gewisse Einflussnahme zu verzeichnen, die darin bestand, die Formulierungen der Berichte sehr detailliert zu korrigieren und eine möglichst positive Berichterstattung zu wünschen. Die Evaluatoren gingen darauf nur im

für sie vertretbaren Rahmen ein und grenzten sich somit von Beeinflussungen ab (vgl. Standard D2). Dass der Auftraggeber sich positive Ergebnisse zu seinem Programm erhoffte, erscheint indes nachvollziehbar und ist nicht als unverhältnismäßiges Interesse dieses Stakeholders zu bewerten (vgl. Abschnitt 5.1.4).

Ein anderer möglicher Interessenkonflikt hätte seitens der Evaluatoren aufkommen können. Denn eine Fortführung der evaluativen Begleitung konnte nur stattfinden, wenn auch das Bundesprogramm verlängert würde, und diese Verlängerung wiederum war wahrscheinlicher, wenn die Evaluation das Programm positiv bewertete. Im vorliegenden Fall stellte dies allerdings keinen Anreiz für die Evaluierenden dar, da zum einen unklar war, ob es überhaupt eine weitere Evaluation geben würde, selbst als die Verlängerung von „Elternchance ist Kinderchance“ feststand. Zum anderen war offen, wer diese mögliche weitere Evaluation durchführen würde, da nicht notwendigerweise das gleiche Evaluationsteam wieder beauftragt werden würde.

Personen mit einem anderen Blickwinkel auf die Evaluation wurden mehrfach einbezogen, indem im Rahmen von Workshops mit der Impulsgruppe ein Austausch mit Experten unterschiedlicher Disziplinen hinsichtlich der Anlage der Evaluation, der Erhebungsinstrumente und der Ergebnisse erfolgte. Dadurch wurden alternative Auffassungen beachtet und zusätzliche Interpretationsmöglichkeiten konnten berücksichtigt werden.

Die Veröffentlichung von Berichten war im Werkvertrag festgelegt, dem zufolge die Entscheidung und Freigabe möglicher Veröffentlichungen dem Auftraggeber oblag. In den Evaluationsberichten wurde versucht, alle Perspektiven abzubilden und die Angaben von Eltern, Fachkräften und weiteren Befragten in fundierten Schlussfolgerungen zu integrieren und damit eine faire, unparteiische Berichterstattung umzusetzen. Bei der Kurzbroschüre scheint dies gelungen zu sein, wie die Stellungnahmen erkennen lassen (vgl. Standard N6). Wie beim vorherigen Standard F3 erwähnt, sind bezüglich des Abschlussberichts noch keine Aussagen zur unparteiischen Berichterstattung bzw. zu deren Einschätzung durch verschiedene Beteiligte möglich, weshalb auch die Bewertung des Standards F4 als vorläufig zu betrachten ist.

Die dargestellten Informationen lassen sich dahingehend zusammenfassen, dass die unterschiedlichen Interessen der Beteiligten zu Beginn der Evaluation erkannt und diese fair behandelt wurden, indem nicht

lediglich einer einzelnen Sichtweise gefolgt wurde. Möglichen Einflussnahmen wurde vorgebeugt bzw. diesen diplomatisch begegnet. Es wurden externe Personen mit alternativen Auffassungen in den Evaluationsprozess miteinbezogen. In den Berichten wurde auf faire und unparteiische Darstellung und die Berücksichtigung aller relevanten Perspektiven geachtet. Da für den Abschlussbericht noch keine diesbezügliche Bewertung der Stakeholder möglich war, ergibt sich folgende (vorläufige) Beurteilung:

⇨ Standard F4 wurde erfüllt.

4.2.2.3.5 Offenlegung der Ergebnisse (F5)

Die Evaluationsergebnisse sollen allen Beteiligten und Betroffenen soweit wie möglich zugänglich gemacht werden.

Die DeGEval-Standards fordern, dass alle Personen, die an einer Evaluation beteiligt oder davon betroffen sind, Zugang zu Ergebnissen und Evaluationsberichten haben. Dafür tragen alle Mitwirkenden an der Evaluation (z. B. Evaluatoren, Auftraggebende, Adressaten) die Verantwortung. Da es häufig sehr viele Beteiligte und Betroffene gibt, ist oftmals die Veröffentlichung des Evaluationsberichts erforderlich, um alle Personen zu erreichen. Die Offenlegung der Ergebnisse sollte in Art und Umfang zu Beginn einer Evaluationsstudie geklärt und im Vertrag festgehalten werden. Ist das nicht möglich, sollte eine Vereinbarung darüber getroffen werden, wer wann auf welche Weise und nach welchen Kriterien über die Offenlegung entscheiden soll. Indem die Ergebnisse einer Evaluation kundgetan werden, soll deren Nutzung gesteigert werden. Jedoch sind auch Fälle denkbar, in denen eine vollständige Offenlegung den Nutzen schmälert, z. B. wenn Konkurrenten eines evaluierten Unternehmens diesem durch die Ergebnisse schaden können, oder wenn die Beteiligung an der Evaluation aufgrund des Wissens um die spätere Publikation (von Schwächen) vermindert wird. In derlei Situationen müssen die Verantwortlichen gemeinsam eine Lösung erarbeiten. Diese kann in einer Beschränkung der Offenlegung bestehen, wobei solche Entscheidungen zu begründen sind.

Die JCSEE-Standards weisen zusätzlich darauf hin, dass nicht nur die Ergebnisse und Empfehlungen der Evaluatoren, sondern ebenso die verwendeten Methoden und die Daten, auf denen die Ergebnisse

basieren, zugänglich gemacht werden sollen, damit diese ggf. kritisch überprüft werden können und die Schlussfolgerungen nachvollziehbar sind.

Da der Abschlussbericht der Evaluation von „Elternchance ist Kinderchance“ veröffentlicht wurde, war der Zugang für alle Beteiligten und Betroffenen gewährleistet, eine Beschränkung der Offenlegung war nicht erforderlich. In diesem umfangreichen Bericht sind neben den gefundenen Ergebnissen und den Schlussfolgerungen die verwendeten Methoden und erhobenen Daten detailliert beschrieben, sodass eine kritische Betrachtung möglich und Nachvollziehbarkeit gegeben ist. Auch die Kurzbroschüre war publiziert worden und ist frei zugänglich (vgl. Abschnitt 4.1.2). Im Evaluationsvertrag war zwar bestimmt, welche Berichte die Evaluatoren zu erbringen hatten und dass sie Inhalte redaktionell aufbereiten sollten, falls der Auftraggeber diese veröffentlichen wollte. Die Offenlegung der Ergebnisse war dort jedoch nicht in Art und Umfang festgehalten, ebenso erfolgte keine Vereinbarung darüber, wer wann und wie eine spätere Bestimmung zur Offenlegung erarbeiten solle. Auch diesbezüglich merkte der Auftraggeber in seiner zweiten Stellungnahme an, dies sei als Merkmal der Auftragsforschung zu sehen. So stelle der Auftraggeber den Nutzer der Ergebnisse dar und dieser entscheide über die weitere Verbreitung.

Da die vertragliche Vereinbarung der Offenlegung der Ergebnisse dient und somit als Grundlage bzw. als möglicher Weg zur Erfüllung dieses Standards betrachtet werden kann, ist das Fehlen einer entsprechenden schriftlichen Vereinbarung im vorliegenden Fall nicht als gravierend einzustufen, da die Ergebnisse letztendlich in vollem Umfang veröffentlicht wurden. Der Standard kann somit als eingehalten bewertet werden, während das Fehlen der entsprechenden vertraglichen Bestimmung in der Diskussion aufgegriffen wird.

⇨ Standard F5 wurde erfüllt.

4.2.2.4 Genauigkeitsstandards

Die Genauigkeitsstandards sollen sicherstellen, dass eine Evaluation gültige Informationen und Ergebnisse zu dem jeweiligen Evaluationsgegenstand und den Evaluationsfragestellungen hervorbringt und vermittelt.

4.2.2.4.1 Beschreibung des Evaluationsgegenstandes (G1)

> Der Evaluationsgegenstand soll klar und genau beschrieben und dokumentiert werden, sodass er eindeutig identifiziert und so genau wie möglich zugänglich gemacht werden kann.

Das Evaluandum ist genau zu beschreiben, damit klar ist, was untersucht wird, so stellen die DeGEval-Richtlinien heraus. Dabei ist zu beachten, dass die Ausgestaltung des Evaluationsgegenstandes in Abhängigkeit vom zeitlichen und räumlichen Kontext unterschiedlich ausfallen kann. Ebenso können zwischen der zu Beginn angedachten und der tatsächlich realisierten Form des Evaluandums Diskrepanzen auftreten. Die genaue Beschreibung des Evaluationsobjektes erleichtert die Zuordnung von Wirkungen und auch Nebenwirkungen zu diesem und ermöglicht außerdem den Vergleich mit anderen Evaluationsgegenständen.

Das JCSEE-Handbuch spricht etwas genauere Empfehlungen aus. So rät es dazu, den Auftraggeber und andere Beteiligte und Betroffene um eine mündliche und, falls möglich, schriftliche Beschreibung der Programmplanung und -wirklichkeit zu bitten. Die Evaluatoren sollten sich nicht mit der Programmbeschreibung des Auftraggebers begnügen, ebenso nicht mit der, welche in der Ausschreibung enthalten ist. Die verfügbaren Programmbeschreibungen sollten hinsichtlich Gemeinsamkeiten und Unterschieden sowie auf Veränderungen im Verlauf untersucht werden. Zudem ist die Genauigkeit der erhaltenen Programmbeschreibungen durch Beobachtung oder durch Befragungen des Programmpersonals zu prüfen. Zu Beginn des Evaluationsprozesses sollten das Programm, das Personal sowie die Teilnehmenden beobachtet werden.

Beim Bundesprogramm „Elternchance ist Kinderchance“ erfolgte die Umsetzung der Maßnahme nicht auf standardisiertem Wege, sondern

gestaltete sich sehr heterogen (vgl. Abschnitt 4.1.1.1). Bereits die Qualifizierung zum Elternbegleiter fand durch zwei Qualifizierungsträger statt, die unterschiedliche Curricula verwendeten. Auch innerhalb eines Trägers gab es eine große Variation hinsichtlich der konkreten Ausgestaltung der Qualifizierungen, da die Curricula den Kursleitern großen Spielraum ließen. Und auch die letztendliche Umsetzung in der Praxis, also die Elternbegleitung selbst, nahm sehr unterschiedliche Formen an, wie sich durch die Evaluation zeigte (vgl. Abschnitt 4.1.4).

Die Heterogenität des Evaluandums erschwerte die Evaluation generell, da eine große Bandbreite an Informationen abgedeckt werden musste und zunächst auch unklar war, wie das neue Bundesprogramm konkret umgesetzt würde. Dementsprechend war auch eine genaue Beschreibung des Evaluationsgegenstandes kaum möglich, da die Untersuchung der Frage, wie die Qualifizierungen genau realisiert wurden und wie die Elternbegleitung in der Praxis aussah, vielmehr Gegenstand der Evaluation selbst war. Es wurde im Abschlussbericht zu Beginn eine möglichst genaue Beschreibung des Bundesprogramms sowie der Qualifizierungen gegeben, einschließlich der Inhalte der beiden Curricula. Die konkrete Umsetzung sowie die zu beobachtende Vielfältigkeit wurden im Ergebnisbericht behandelt. Somit kann unter Berücksichtigung der Umstände von einer genauen Beschreibung des Evaluationsgegenstandes ausgegangen werden.

Eine Beobachtung des Programms einschließlich des Personals und der Teilnehmenden war durch zwei Mitglieder des Evaluationsteams vorgenommen worden, welche für einige Tage bei einem Qualifizierungskurs hospitierten und dadurch hilfreiche Eindrücke gewannen. Das Evaluationsteam hatte im Rahmen der Ausschreibung eine schriftliche Programmbeschreibung des Auftraggebers erhalten. Beschreibungen der Programmplanung und -wirklichkeit von anderen Beteiligten waren nicht schriftlich eingeholt worden (mit Ausnahme der beiden Weiterbildungscurricula der Träger), sondern wurden im Rahmen des persönlichen Austausches und der Experteninterviews mündlich erfasst. Dies galt ebenso für spätere Programmauffassungen des Auftraggebenden. Die von den Standards geforderte Untersuchung der Genauigkeit dieser Beschreibungen, von Gemeinsamkeiten und Unterschieden sowie Veränderungen im Verlauf gestaltete sich aufgrund der beschriebenen Heterogenität des Programms als schwierig, da kein Abgleich mit einem konkreten, „richtigen“ Vorgehen möglich und auch nicht sinnvoll war.

Insgesamt zeigt sich, dass dieser Standard nur begrenzt auf die vorliegende Evaluation anwendbar ist bzw. dass zwar Abstriche zu verzeichnen sind, diese jedoch in der Beschaffenheit des Evaluationsgegenstandes begründet liegen. Da sich die Evaluatoren dennoch um eine genaue Beschreibung bemühten sowie die weiteren Anforderungen des Standards nach Möglichkeit einhielten, wird der Standard als eingehalten bewertet.

⇨ Standard G1 wurde erfüllt.

4.2.2.4.2 Kontextanalyse (G2)

Der Kontext des Evaluationsgegenstandes soll ausreichend detailliert untersucht und analysiert werden.

In den DeGEval-Richtlinien ist hierzu vermerkt, dass der Kontext des Evaluandums sich aus dessen Rahmenbedingungen zusammensetzt, worunter z. B. soziale und politische Einflüsse, die Interessen und Eigenschaften der maßgeblich Beteiligten und Betroffenen oder auch wirtschaftliche Begleitumstände fallen können. Eine fundierte Kontextanalyse ist wichtig, weil sie Informationen über mögliche Bedingungsfaktoren erbringen kann, welche den Evaluationsgegenstand und die Evaluationsergebnisse beeinflussen. Des Weiteren können sich daraus Erkenntnisse in Bezug auf die Generalisierbarkeit der Evaluationsbefunde ableiten lassen. Der Kontext sollte nicht zu eng definiert, die Kontextanalyse aber auch nicht zu breit durchgeführt werden, damit die dazu notwendigen Ressourcen nicht bei anderen Evaluationsschritten fehlen.

Das JCSEE-Handbuch ergänzt den Aspekt der Untersuchung der Kontextfaktoren um den der Beschreibung derselben. Die Erläuterungen zum Kontext sollen die Beteiligten und Betroffenen demnach dabei unterstützen, die Vergleichbarkeit der vorliegenden Rahmenbedingungen mit anderen Kontexten abzuschätzen.

Da bei der Evaluation von „Elternchance ist Kinderchance“ unter anderem untersucht werden sollte, wie sich die Elternbegleitung in der Praxis gestaltete, wurden einige Kontextfaktoren direkt erfragt (z. B. Einrichtungsart, Leitungsposition der Fachkraft), während weitere

offen ermittelt wurden, indem beispielsweise nach Erleichterungen und Erschwernissen der Arbeit der Fachkräfte gefragt wurde. Insgesamt konnten dadurch etliche Einflussgrößen identifiziert werden. Diese wurden bei den Berechnungen und Interpretationen einbezogen, hinsichtlich ihrer Bedeutung für die Generalisierbarkeit der Ergebnisse abgewogen und im Abschlussbericht beschrieben. Des Weiteren war auch der politische bzw. wirtschaftliche Kontext zu berücksichtigen, in dem das Bundesprogramm stattfand. So war es z. B. aufgrund der Art der dem BMFSFJ zur Verfügung stehenden Mittel zwar möglich, die Qualifizierungen zu Elternbegleitern zu finanzieren, jedoch konnten und sollten aus diesen Mitteln keine Arbeitsstellen geschaffen werden, um dadurch dem von den Befragten kritisierten Zeit- und Ressourcenmangel Abhilfe zu schaffen. Dieser Umstand wurde im Abschlussbericht erwähnt.

Es wurden somit viele Kontextfaktoren ermittelt, analysiert und im Abschlussbericht beschrieben. Der dazu benötigte Aufwand kann als angemessen beurteilt werden. Die Kontextanalyse und -beschreibung ist somit gelungen und in angemessenem Umfang erfolgt, sodass folgende Bewertung ergeht:

⇨ Standard G2 wurde erfüllt.

4.2.2.4.3 Beschreibung von Zwecken und Vorgehen (G3)

Gegenstand, Zwecke, Fragestellungen und Vorgehen der Evaluation, einschließlich der angewandten Methoden, sollen genau dokumentiert und beschrieben werden, sodass sie identifiziert und eingeschätzt werden können.

Laut DeGEval-Standards sollen die Evaluationszwecke und Evaluationsfragestellungen im Evaluationsverlauf genau dokumentiert und im Zuge der Berichterstattung nachvollziehbar erläutert werden. Dabei sind auch abweichende Sichtweisen zu berücksichtigen. Ebenso sollen das Vorgehen, die Methoden und die Entscheidungen zum Evaluationsplan, einschließlich ihrer Begründungen, im Verlauf aufgezeichnet und im Bericht dargelegt werden. So sind die organisatorischen Vorkehrungen, die Erhebungen (ggf. einschließlich Stichprobenverfahren), die Aufbereitung, Analyse und Interpretation der Daten sowie die

Berichterstattung detailliert zu dokumentieren. Dabei ist zu beachten, dass das Vorgehen im Evaluationsverlauf geändert bzw. angepasst werden kann und die tatsächliche Umsetzung sich evtl. nicht mit der geplanten deckt. Abweichungen vom geplanten Vorgehen sollen deutlich gemacht und begründet werden. Auch der Zeitplan und Abweichungen davon sollen schriftlich fixiert werden.

Die Angaben im JCSEE-Handbuch stimmen weitestgehend mit denen der DeGEval überein. Ergänzend wird dort erklärt, welche Zielsetzung dieser Standard verfolgt. So ist eine detaillierte und begründete Beschreibung des Vorgehens einer angemessenen Bewertung der Ergebnisse und Empfehlungen durch die Beteiligten dienlich. Sie kann zudem die Informationsgrundlage für eine Meta-Evaluation darstellen und auch als Anhaltspunkt für ähnliche Evaluationen dienen. Weiterhin wird damit eine Orientierung für Wiederholungen der Evaluation geschaffen. Und schließlich können aus solchen detailreichen Schilderungen Fallbeispiele zur Nutzung in der Ausbildung von Evaluatoren generiert werden.

Die Zwecke und Fragestellungen der vorliegenden Evaluation wurden zu Beginn in der Ausschreibung und im Vertrag schriftlich festgehalten, im Verlauf in Form von Sitzungsprotokollen dokumentiert und auch in Abschlussbericht und Kurzbroschüre dargestellt. Sie wurden zudem unter anderem durch Veröffentlichung auf der Homepage des Bundesprogramms frühzeitig an die Beteiligten kommuniziert. Auch Informationen zum Vorgehen, den geplanten Methoden und zum Zeitplan waren anfangs in Ausschreibung, Projektangebot und Werkvertrag schriftlich festgehalten worden. Die tatsächliche Ausführung sowie notwendige Anpassungen wurden eng mit dem Auftraggeber abgestimmt, protokolliert und im Abschlussbericht offengelegt. Beispielsweise erwies sich die Erhebung der Eltern-Kontrollgruppe aufgrund einer sehr geringen Teilnehmerquote als äußerst aufwendig und konnte erst deutlich später als geplant abgeschlossen werden, sodass diese Daten nach einer entsprechenden Vereinbarung in einem zusätzlichen Bericht nachgeliefert wurden.

Eine Abweichung vom ursprünglich geplanten Vorgehen, die nicht im Evaluationsbericht festgehalten wurde, bestand in dem Wunsch des Auftraggebers nach vergleichenden Auswertungen der Curricula der beiden Träger. Diese stellten eine Änderung der Evaluationszwecke dar, sollten jedoch im Abschlussbericht nicht erwähnt werden. Dort

waren lediglich trägerübergreifende Ergebnisse enthalten. Eine offenere Kommunikation dieser Abweichung im Evaluationsverlauf den Trägern gegenüber wäre wünschenswert gewesen (vgl. Standard N2).

Zusammengenommen wurden die Evaluationszwecke sowie -fragestellungen sowie das Vorgehen einschließlich Methoden und Zeitplan zu Beginn der Evaluation und im Verlauf dokumentiert und im Abschlussbericht dargelegt. Da jedoch nicht alle Abweichungen vom ursprünglich geplanten Vorgehen berichtet bzw. den Betroffenen kommuniziert wurden, ergibt sich nachstehende Bewertung:

⇨ Standard G3 wurde überwiegend erfüllt.

4.2.2.4.4 Angabe von Informationsquellen (G4)

Die im Rahmen einer Evaluation genutzten Informationsquellen sollen hinreichend genau dokumentiert werden, damit die Verlässlichkeit und Angemessenheit der Information eingeschätzt werden kann.

Bei Evaluationsstudien kann eine Vielzahl von Informationsquellen zur Anwendung kommen, erläutern die DeGEval-Standards. Dies können z. B. Individuen oder Gruppen sein, aber auch Dokumente, Statistiken oder audiovisuelle Dateien können herangezogen werden. Der Einbezug unterschiedlicher Informationsquellen lässt einen Vergleich der erhobenen Daten zu. Diese Daten sollen eingeschätzt und bewertet werden können. Die verwendeten Informationsquellen sind daher möglichst genau zu beschreiben, damit die Adressaten die Qualität der gewonnenen Informationen selbst beurteilen können. Eine fehlende oder unzureichende Darstellung der Informationsquellen kann hingegen die Glaubwürdigkeit der Evaluation beeinträchtigen.

Im Handbuch des JCSEE wird weiterhin darauf eingegangen, dass Evaluierende in der Regel nur einen Teil der potentiell verfügbaren Informationen erhalten, da nicht jede Person der interessierenden Grundgesamtheit befragt werden kann. Daher müssen Stichproben gezogen werden, die jedoch im Evaluationsverlauf Veränderungen unterliegen können. Sowohl die Kriterien für die Stichprobengewinnung als auch etwaige Veränderungen sind von den Evaluatoren zu dokumentieren. Die Erhebungsverfahren, die bei den verschiedenen

Informationsquellen zur Anwendung kommen, sollten ebenfalls dokumentiert sowie Kopien der Instrumente aufbewahrt und in den Anhang des Evaluationsberichts aufgenommen werden.

Bei der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ wurden bei einer Vielzahl von Informationsquellen (Eltern, Fachkräfte, Kursleiter, Programmverantwortliche, Einrichtungsleiter sowie Kooperationspartner von EB-Plus-Einrichtungen) anhand verschiedener Methoden (quantitative Erhebung mittels Fragebögen sowie qualitative Befragung durch Tiefeninterviews und Fokusgruppen) Daten erhoben, welche aufeinander bezogen und verglichen werden konnten. An den Methoden kann kritisiert werden, dass diese zwar verschiedene Ansätze einbezogen, die damit erfassten Informationen aber dennoch lediglich auf Selbstberichten beruhten. Tatsächlich war im Modul 1 der Evaluation ursprünglich auch der Einsatz von Videovignetten geplant gewesen, um den Kompetenzzuwachs bei den Fachkräften objektiver messen zu können. Allerdings stellte sich heraus, dass die Fachkräfte in unterschiedlichsten Arbeitsfeldern tätig waren und entsprechend eine Vielzahl verschiedener Elternkontakte erlebten (z. B. Erzieher in Kindertagesstätten vs. Erwachsenenbildner in Volkshochschulen). Dies erschwerte die Generierung eines einheitlichen Messinstruments in Form einer Videosequenz, die für jede Fachkraft geeignet war, sehr. Hinzu kam die Tatsache, dass die Qualifizierungskurse keineswegs einheitlich umgesetzt wurden, sodass Elternbegleitern aus verschiedenen Kursen nicht unbedingt die gleichen Inhalte vermittelt worden waren. Die Videovignetten konnten somit aufgrund der Heterogenität sowohl der Tätigkeiten der Fachkräfte als auch der Qualifizierungskurse sowie aufgrund zeitlicher und finanzieller Begrenzungen nicht realisiert werden.

Im Abschlussbericht wurde die Methodik der Evaluation genau beschrieben. Dies umfasste das Evaluationsdesign, die Datengewinnung und die Erhebungsinstrumente. Im Rahmen der Datengewinnung wurden der Zugang zu den Stichproben beschrieben sowie mögliche Verzerrungen derselben durch Drop-Outs oder positive Selektion thematisiert. Auch die Stichprobe selbst wurde detailliert vorgestellt. Im Hinblick auf die Erhebungsinstrumente wurde auf eine Aufnahme der Fragebögen in den Anhang des Berichts verzichtet, da dieser sonst mehrere hundert Seiten umfasst hätte. Die verwendeten Instrumente

wurden dafür im Methodenteil des Berichts möglichst genau geschildert.

Alles in allem wurden die Informationsquellen der vorliegenden Evaluation im Abschlussbericht genau beschrieben und dem Leser somit detaillierte Informationen zur Verfügung gestellt, um sich eine eigene Vorstellung von der angewendeten Methodik machen und diese kritisch bewerten zu können. Der Standard kann damit als eingehalten gelten.

⇨ Standard G4 wurde erfüllt.

4.2.2.4.5 Valide und reliable Informationen (G5)

> Die Verfahren zur Gewinnung von Daten sollen so gewählt oder entwickelt und dann eingesetzt werden, dass die Zuverlässigkeit der gewonnenen Daten und ihre Gültigkeit bezogen auf die Beantwortung der Evaluationsfragestellungen nach fachlichen Maßstäben sichergestellt sind. Die fachlichen Maßstäbe sollen sich an den Gütekriterien quantitativer und qualitativer Sozialforschung orientieren.

Die DeGEval-Standards fordern, dass bei der Datenerhebung im Rahmen einer Evaluation anerkannte Gütekriterien eingehalten werden, damit die erfassten Daten möglichst wenig fehlerbehaftet sind. Quantitative Verfahren und die mit ihnen gewonnenen Informationen sollen demnach objektiv (unabhängig vom Testleiter), reliabel (messgenau, d. h. möglichst wenig anfällig für Störeinflüsse oder zufällige Fehler) und valide (inhaltlich passend, d. h. es wird das gemessen, was gemessen werden soll) sein. Bei qualitativen Zugängen und nicht-standardisierten Instrumenten sind entsprechende Gütekriterien der qualitativen Forschung anzulegen, darunter beispielsweise die Triangulation, bei welcher unterschiedliche Datenquellen verbunden werden, um zu einer geeigneten Interpretation zu kommen. Hinsichtlich der Validität wird zudem erläutert, dass diese sich nicht nur auf die Qualität der Erhebungsinstrumente und Daten bezieht, sondern auch die Güte und Glaubwürdigkeit der Schlüsse meint, die aus den Ergebnissen gezogen werden. Demnach ist Validität nur im jeweiligen Evaluationskontext und unter Bezug auf die Zielsetzung und Fragestellungen

einzuschätzen. Die Interpretationen sind durch Belege zu stützen. Zur Validierung werden Belege zusammengetragen, weshalb verschiedene Datenerhebungsverfahren eingesetzt werden sollten. Insgesamt kommt den genannten wissenschaftlichen Gütekriterien eine Orientierungsfunktion hinsichtlich Auswahl, Entwicklung und Anwendung von Methoden zu und sie sind den Anforderungen der Evaluationspraxis entsprechend zu spezifizieren und zu ergänzen.

Im JCSEE-Handbuch ist dieser Standard auf zwei Richtlinien aufgeteilt, nämlich „Valide Informationen“ und „Reliable Informationen“. Bei den Ausführungen zur Validität wird detaillierter erläutert, warum die Anwendung verschiedener Verfahren und Informationsquellen anzustreben ist. Dadurch kann einerseits sichergestellt werden, dass alle wichtigen Variablen beachtet werden. Andererseits können alle Informationsquellen fehlerbehaftet sein, sodass bei einer Kombination von Verfahren die Schwächen eines Ansatzes durch die Stärken eines anderen ausgeglichen werden können. Allerdings muss der durch den Einsatz vieler Verfahren erreichbare Informationsgewinn gegen die Belastung der Befragten abgewogen werden. Das Handbuch gibt weiterhin den Hinweis, dass die Passung zwischen den Erhebungsverfahren und den Zielen und Inhalten des Evaluandums geprüft werden sollte und dabei relevante Beteiligte und Betroffene einzubeziehen sind. Die Wahl der einzelnen Verfahren sollte im Methodenteil des Evaluationsberichts begründet werden. Bei neu entwickelten Verfahren soll deren Experimentalcharakter aufgezeigt und zudem erwähnt werden, dass die Ergebnisse vorsichtig interpretiert werden müssen. Weiterhin sollten die Personen, die ein Verfahren durchführen, dazu qualifiziert und ausreichend vorbereitet sein. Seitens der Befragten ist auf solche Eigenschaften zu achten, welche die Validität einschränken könnten, darunter z. B. Lesefähigkeit oder sprachliches Ausdrucksvermögen. Hinsichtlich der Reliabilität wird im Handbuch erläutert, dass Abweichungen, die durch systematische Faktoren im Zusammenhang mit der evaluierten Maßnahme hervorgerufen werden, von solchen zu unterscheiden sind, die auf Störeinflüsse oder Zufallsfehler zurückgehen. Sind Evaluationsergebnisse nicht mit Faktoren des Evaluandums erklärbar, stellt dies die Reliabilität in Frage. Evaluatoren sollten Aussagen zur Reliabilität ihrer Verfahren treffen, auch bei neu entwickelten Instrumenten.

Zur Einhaltung des Gütekriteriums der Objektivität wurden in der vorliegenden Evaluation verschiedene Maßnahmen getroffen. So erfolgten die quantitativen Erhebungen in den Modulen 1 und 2 anhand standardisierter Online-Fragebögen. Im Modul 3 wurde der Zugang der telefonischen Befragung gewählt, die Fragebögen waren jedoch auch hier standardisiert und die Interviewer wurden umfassend geschult. Die Durchführung der qualitativen Interviews in allen Modulen erfolgte ebenfalls durch erfahrene bzw. eigens geschulte Personen. Bei der Kodierung des qualitativen Materials wurden im Modul 3 mehrere Beurteiler eingesetzt und die Übereinstimmungen berechnet, um Objektivität und Reliabilität sicherzustellen. In den beiden anderen Modulen war dies nicht der Fall, da hier andere Auswertungsansätze verfolgt wurden.

Hinsichtlich der Reliabilität wurden für die Skalen der quantitativen Befragung des Moduls 1 die internen Konsistenzen berechnet und im Anhang des Abschlussberichts dargestellt. In den Modulen 2 und 3 lagen kaum Skalen vor, daher erfolgte keine Reliabilitätsbestimmung.

Im Hinblick auf die Validität ist festzuhalten, dass bei mehreren Informationsquellen und mittels verschiedener Methoden Daten erhoben wurden (die allerdings sämtlich auf Selbstberichten der Befragten beruhten, vgl. hierzu Standard G4). Diese Daten konnten abgeglichen und aufeinander bezogen werden, sodass eine Validierung und Triangulation möglich war. Zudem wurden die Vollständigkeit und die inhaltliche Passung zwischen den Erhebungsinstrumenten und den Zielen der Evaluation zu Beginn intensiv mit dem Auftraggeber abgestimmt. Auch die Gruppen der Elternbegleiter und Eltern wurden im Rahmen der Pilotstudien einbezogen. Ergänzend hätte noch die Meinung von Elternbegleitern zum Elternfragebogen eingeholt werden können, wie in einigen Stellungnahmen angemerkt wurde (vgl. Standard N1).

Generell kam bei der vorliegenden Evaluation eine Kombination aus bestehenden, anerkannten Erhebungsinstrumenten und neu entwickelten Items bzw. Skalen zum Einsatz. Hierin zeigte sich die Notwendigkeit der Anpassung an die Anforderungen der Evaluation, da einerseits nicht für alle interessierenden Bereiche Verfahren vorlagen und somit einige Instrumente neu konstruiert werden mussten. Andererseits konnte von existierenden Skalen zum Teil nur eine eigens konstruierte Kurzform verwendet werden, um die Probanden zeitlich nicht zu sehr zu belasten (vgl. Standard D1). Die Erhebungsinstrumente

wurden im Methodenteil des Abschlussberichts dargestellt, ihre Wahl wurde jedoch nicht im Einzelnen begründet, um diesen Abschnitt und den Bericht im Ganzen nicht noch länger werden zu lassen. Dabei wurde benannt, welche Items bzw. Skalen neu entwickelt worden waren, auf die sich daraus ergebenden möglichen Einschränkungen der Validität wurde hingegen nicht explizit hingewiesen.

Seitens der Probanden waren in den Modulen 1 und 2, in welchen Fachkräfte und Einrichtungsleitungen befragt wurden, keine validitätsgefährdenden Einschränkungen zu erwarten. In Modul 3 hingegen wurde aufgrund der Zielgruppen des Bundesprogramms damit gerechnet, dass vereinzelt Schwierigkeiten beim sprachlichen Verständnis bzw. Ausdruck auftreten könnten. Die Interviewer sollten daher nach jedem Interview ihren Eindruck vom Sprachniveau des Probanden angeben und die Fälle notieren, bei denen das Verständnis des Fragebogens ihrer Einschätzung nach nicht vollständig gegeben war. Diese wurden von den Analysen ausgeschlossen.

Hinsichtlich der Validität ist zudem zu beachten, dass bei den Probanden des Moduls 3 nicht, wie ursprünglich geplant, eine Zufallsstichprobe aus vermittelten Eltern gezogen werden konnte, weil die Zahl der von Fachkräften vermittelten Personen zu gering ausfiel. Diese hätte nur durch beträchtlichen Mehraufwand verbessert werden können, was allerdings stark zu Lasten der Durchführbarkeit der Evaluation gegangen wäre und daher nicht umgesetzt werden konnte (vgl. Standard D1). Die resultierende Elternstichprobe war somit einem Selektionsprozess unterzogen, der sich nicht auf die Selbstselektion im Sinne der Freiwilligkeit der Teilnahme beschränkte (diese Art der Selektion traf auf alle Befragten bzw. Stichproben der Evaluation zu), sondern auch eine vermutlich selektive Vermittlung von Eltern durch Fachkräfte umfasste. Diese Stichprobe konnte daher nicht als repräsentativ für alle von Fachkräften begleiteten Eltern gelten, was die Generalisierbarkeit der Ergebnisse einschränkte.

Als Fazit ergibt sich, dass die Evaluation die Gütekriterien der qualitativen Forschung gut erfüllte. Im quantitativen Bereich der Studie wurde den Kriterien der Objektivität und Validität weitgehend entsprochen, wohingegen die Reliabilität als nicht hinreichend überprüft einzustufen ist. Deshalb wird folgende Bewertung vorgenommen:

⇨ Standard G5 wurde teilweise erfüllt.

4.2.2.4.6 Systematische Fehlerprüfung (G6)

> Die in einer Evaluation gesammelten, aufbereiteten, analysierten und präsentierten Informationen sollen systematisch auf Fehler geprüft werden.

Laut DeGEval-Standards können bei den Schritten der Datenerhebung, -aufbereitung, -auswertung und -interpretation sowie bei der Präsentation von Ergebnissen verschiedenste Fehler auftreten. Dies können methodische Fehler sein, z. B. Mängel im Design durch unzureichende Kontrolle von Störvariablen, fehlerhafte Stichprobenziehung oder auch Missachtung der Voraussetzungen statistischer Verfahren. Andere Fehler können aufgrund mangelnder Sorgfalt entstehen, beispielsweise wenn Probanden-Codes verwechselt, Daten falsch eingegeben oder Beschriftungen in Präsentationen vertauscht werden. Fehler können inkorrekte Interpretationen nach sich ziehen und die Evaluation damit in Verruf bringen. Daher ist im Evaluationsprozess dafür zu sorgen, dass potentielle Fehlerquellen frühzeitig erkannt und Fehler nach Möglichkeit vermieden oder korrigiert werden. Dies kann durch Schulungen der beteiligten Personen und durch systematische Kontrollen und Genauigkeitsprüfungen erreicht werden. Mögliche Fehler und deren Auswirkungen sollten berichtet werden.

Das JCSEE-Handbuch weist weiterhin darauf hin, dass die Evaluatoren auch bei hochqualifizierten und engagierten Fachkräften nicht davon ausgehen sollten, dass diese fehlerfrei arbeiten oder ihre eigene Arbeit überprüfen. Das Evaluationspersonal sollte ausreichend vorbereitet und zudem instruiert sein, auf mögliche typische Fehler zu achten. Zudem dürfen die Evaluierenden sich nicht darauf verlassen, dass die Befragten alle Anweisungen lesen, sie verstehen und befolgen.

Bei der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ wurden viele Maßnahmen ergriffen, um möglichst fehlerfreie Daten zu erhalten. Dabei ist zunächst die Pilotstudie zu nennen, bei der in allen drei Modulen die Fragebögen auf Verständlichkeit hin geprüft wurden. Die qualitativen Erhebungen erfolgten in allen Modulen durch geschulte Interviewer (vgl. Standard G5). Für die quantitative Erhebung im Modul 3 wurden die Interviewer ebenfalls geschult und darauf hingewiesen, dass sie mögliche Vorkommnisse sowie fehlerhafte Eingaben sofort notieren und dem Evaluationsteam rückmelden

sollten. Dadurch konnten beispielsweise Fehler bei der Eingabe der Probanden-Codes direkt ausfindig gemacht und korrigiert werden. Zudem ermöglichte dieses Vorgehen, aufgrund der gesammelten Erfahrungen mit häufigen Fehlern, eine noch bessere Vorbereitung der neu hinzukommenden Interviewer.

Dass die Instruktionen der Evaluatoren nicht immer genau beachtet werden, war auch im vorliegenden Projekt zu beobachten. So waren beispielsweise die Fachkräfte, welche dem Evaluationsteam Eltern vermittelten, explizit darum gebeten worden, nur Eltern zu vermitteln, mit denen sie zusammenarbeiteten. Bei der Kontaktierung der Eltern stellte sich jedoch heraus, dass dies nicht immer der Fall war. Dieser Fehlerquelle wurde entgegengewirkt, indem Art und Dauer des Kontakts in der Erhebung erfasst und bei der Auswertung kontrolliert wurden.

Die erhobenen Daten wurden vor ihrer Analyse einer genauen Kontrolle unterzogen, indem sie z. B. auf inkonsistente oder unlogische Angaben und Ausreißer geprüft wurden. Dabei konnten einige Fehler aufgedeckt und in den meisten Fällen korrigiert werden. Wenn dies nicht möglich war, wurde die entsprechende Variable mit einem fehlenden Wert belegt und in ganz seltenen Fällen der Proband von den Analysen ausgeschlossen, wenn der fragliche Wert darauf hindeutete, dass er nicht zur interessierenden Population gehörte (z. B. das Kind nicht im richtigen Alter war). Bei den Analysen selbst wurden die Anwendungsvoraussetzungen gründlich geprüft und die angemessenen statistischen Verfahren verwendet.

Da Fehler auf Variablenebene somit bereits vor der Analyse und Interpretation behoben wurden, wurden diese nicht im Einzelnen im Abschlussbericht aufgeführt. Berichtet wurden dort hingegen mögliche Verzerrungen der Stichproben sowie die denkbaren Konsequenzen für die Übertragbarkeit der Ergebnisse.

Zusammengenommen wurden viele Bemühungen unternommen, um Fehler in den verschiedenen Evaluationsphasen zu vermeiden, die erhaltenen Daten zu überprüfen und ggf. zu korrigieren. Der Standard kann damit als eingehalten bewertet werden.

⇨ Standard G6 wurde erfüllt.

4.2.2.4.7 Analyse qualitativer und quantitativer Informationen (G7)

> Qualitative und quantitative Informationen einer Evaluation sollen nach fachlichen Maßstäben angemessen und systematisch analysiert werden, damit die Fragestellungen der Evaluation effektiv beantwortet werden können.

Die Datenanalyse bildet die Grundlage für Interpretationen und Schlussfolgerungen im Hinblick auf die Fragestellungen einer Evaluation. Den DeGEval-Richtlinien zufolge sind die Erhebungs- und Analyseverfahren den Fragestellungen, dem Kenntnisstand bezüglich des Evaluandums und den Kontextfaktoren des Evaluationsfeldes entsprechend auszuwählen. Vorlieben der auswertenden Personen hingegen sollten die Auswahl nicht beeinflussen. Die Datenauswertung soll systematisch durchgeführt und entsprechende Regeln für qualitative sowie für quantitative Verfahren sollen beachtet werden. Die herangezogenen Regeln und ihre methodischen Grundlagen sind, beispielsweise durch Literaturnachweise, offenzulegen. Auswahl und Anwendung der Instrumente sollen transparent und nachvollziehbar erfolgen, um eine Kritik an den Auswahlentscheidungen sowie den Ergebnissen zu ermöglichen. Kennwerte und Formeln sind allgemein verständlich zu erläutern. Die Aussagekraft der Methoden sowie deren Einschränkungen sind darzulegen.

Auch in diesem Standard wurden zwei Richtlinien des JCSEE zusammengefasst, nämlich „Analyse quantitativer Informationen“ und „Analyse qualitativer Informationen“. Das Handbuch empfiehlt für die Auswertung quantitativer Daten, dass die Evaluatoren mit vorbereitenden Probeanalysen beginnen, um sich mit den Daten vertraut zu machen, und nachfolgend tiefergehende Auswertungen vornehmen, die belastbare Ergebnisse erbringen. Dabei sind die praktische Relevanz (z. B. Effektstärken) und die statistische Signifikanz zu beachten. Die Ergebnisse sollten graphisch aufbereitet und im Bericht veranschaulicht werden. Generell sollte die Komplexität der Analysen auch den Bedürfnissen der Adressaten angepasst werden, denen in manchen Fällen mit einfacheren Analysen besser gedient sein könnte. Bei vergleichenden Berechnungen mit Gruppen, die nicht per Zufall entstanden sind, sollten die Gruppenunterschiede statistisch kontrolliert oder zumindest unterschiedliche Auswertungsmethoden angewendet werden, um die Stabilität der Ergebnisse einschätzen zu

können. Wenn Untergruppen ausgemacht werden, sollten die Analysen diese berücksichtigen. Das Handbuch führt weiterhin aus, dass quantitative und qualitative Auswertungen einander ergänzen und dass Interpretationen sowie Empfehlungen auf beiden basieren sollten. Auch die qualitativen Daten sind angemessen auszuwerten, wobei in der Regel die Bildung von Kategorien eine zentrale Rolle spielt, welche für die Beantwortung der Evaluationsfragestellungen geeignet sind und in die das Datenmaterial eingeordnet werden kann. Diese Kategorien sind auf Reliabilität und Validität hin zu prüfen, indem z. B. zwei unabhängige Beurteiler das gleiche Material einordnen oder eine externe Person die Einheitlichkeit der Analysen untersucht. Den Beteiligten und Betroffenen sollten vorläufige Auswertungen, Ergebnisse und Empfehlungen vorgestellt werden, um sicherzugehen, dass sie diese für geeignet bzw. sinnvoll erachten.

Im Rahmen der vorliegenden Evaluation kamen verschiedene quantitative und qualitative Analysemethoden zum Einsatz, um das Datenmaterial systematisch auszuwerten. Die Auswahl der Auswertungsverfahren orientierte sich an den Fragestellungen der Evaluation sowie am Datenmaterial selbst und wurde einschließlich Literaturnachweisen im Abschlussbericht dargelegt. Die Erhebungsinstrumente waren an den Kenntnisstand zum Evaluandum und an Kontextfaktoren angepasst. So war z. B. noch unklar, wie sich die Elternbegleitung in der Praxis gestaltete, sodass die Fragebögen diesen Bereich ausführlich erfassten. Der Kontext wurde beispielsweise dahingehend berücksichtigt, dass die quantitative Erhebung bei den EB-Plus-Einrichtungen so kurz wie möglich gehalten wurde, da die Einrichtungsleitungen sehr eingebunden waren und zudem auch die Servicestelle dort Daten erhob.

Bei den quantitativen Analysen wurden zunächst einfachere Auswertungen durchgeführt, anschließend erfolgten komplexere, tiefergehende Analysen, wobei die Anwendungsvoraussetzungen stets geprüft wurden. Dabei wurden auch Subgruppen berücksichtigt. Interventions- und Kontrollgruppen wurden auf Unterschiede hin untersucht, diese wurden berichtet sowie nach Möglichkeit statistisch kontrolliert. Signifikanzen und Effektstärken wurden beachtet und auch im Abschlussbericht dargestellt und diskutiert. Dort wurde ebenfalls eine Diskussion der Aussagekraft der Methoden sowie von deren Einschränkungen vorgenommen.

Die Darstellung der Ergebnisse erfolgte im Abschlussbericht ebenso wie in der Kurzbroschüre anschaulich unter Verwendung von Graphiken. In der Kurzbroschüre wurde im Hinblick auf die damit zu erreichende Zielgruppe auf den Bericht komplexer Analysen und statistischer Kennwerte verzichtet und mehr Wert auf Anschaulichkeit und Verständlichkeit gelegt. Der Abschlussbericht enthielt demgegenüber alle Analysen, wobei grundlegende Kennwerte wie z. B. Signifikanzniveaus nicht erklärt wurden, da der Bericht auf ein eher wissenschaftliches Publikum ausgerichtet war. Insofern waren die Komplexität der Analysen sowie die Art der Darstellung durchaus an die Bedarfe der Adressaten angepasst. Dennoch stellte es sich als Herausforderung dar, die Ergebnisse der nach Ansicht des Evaluationsteams angemessenen, komplexen Auswertungen so zu vereinfachen, dass diese dem Bedürfnis des Auftraggebers nach verständlicher Erklärung entsprachen und dennoch die Befunde korrekt abgebildet wurden. Dies war z. B. der Fall bei signifikanten Dreifachinteraktionen im Rahmen von Varianzanalysen im Modul 3, welche sich bei Einbezug der Subgruppen ergaben und nicht mit einfachen Worten erklärbar waren. Hier bedurfte es einiger Formulierungsvorschläge seitens der Evaluatoren, bis ein Kompromiss gefunden war, der beide Stakeholder zufriedenstellte.

Erste Ergebnisse und Empfehlungen aus den qualitativen Daten sowie das Vorgehen bei der Auswertung wurden verschiedenen Beteiligten (Auftraggeber, Kompetenzteam und Impulsgruppe) vorgestellt und deren Meinungen dazu eingeholt. Insgesamt ergänzten sich qualitative und quantitative Informationen und erbrachten gemeinsam wertvolle Befunde und Schlussfolgerungen.

Alles in allem wurden die quantitativen und qualitativen Daten systematisch und fachlichen Maßstäben entsprechend ausgewertet. Die sorgfältigen Analysen lieferten informative Ergebnisse und ermöglichten eine umfassende Beantwortung der Evaluationsfragestellungen. Daher ergibt sich folgende Bewertung:

⇨ Standard G7 wurde erfüllt.

4.2.2.4.8 Begründete Schlussfolgerungen (G8)

> Die in einer Evaluation gezogenen Folgerungen sollen ausdrücklich begründet werden, damit die Adressatinnen und Adressaten diese einschätzen können.

Wie den DeGEval-Standards zu entnehmen ist, sollen die Schlussfolgerungen, die sich im Rahmen einer Evaluationsstudie ergeben, explizit begründet werden. Sie sollen zusammen mit den zugrunde liegenden Annahmen und den angewendeten Verfahren transparent dargestellt werden, wobei auch der Geltungsbereich der Schlüsse deutlich zu machen ist. Sofern es erforderlich ist, sollen im Bericht auch alternative Interpretationen diskutiert werden, einschließlich einer Begründung, warum diese verworfen wurden. Wird den Forderungen dieses Standards entsprochen, so stärkt das die Überzeugungskraft der Schlussfolgerungen und ermöglicht weiterhin den Nutzern der Ergebnisse, deren Aussagekraft selbst einzuschätzen.

Die Ausführungen des JCSEE-Handbuchs zu diesem Standard sind zwar etwas ausführlicher als die der DeGEval, jedoch stimmen die zentralen Aspekte überein, sodass an dieser Stelle keine Hinweise aus dem JCSEE-Handbuch zu ergänzen sind.

Im Abschlussbericht der Evaluation von „Elternchance ist Kinderchance“ wurde den Anforderungen dieses Standards nachgekommen. Im Methodenteil wurden das Evaluationsmodell einschließlich der grundlegenden Annahmen sowie die Erhebungsinstrumente und deren Anwendung beschrieben. Die Schlussfolgerungen basierten klar auf den erhobenen Daten und waren somit begründet. Mögliche Einschränkungen des Geltungsbereichs wurden ebenfalls dargelegt. Im Hinblick auf alternative Interpretationsmöglichkeiten ist anzumerken, dass die Studie so angelegt war, dass durch das Evaluationsdesign vielen Alternativerklärungen vorgebeugt wurde. So wurden Kontrollgruppen einbezogen, damit Veränderungen nicht beispielsweise auf rein zeitliche Entwicklungseffekte während der Dauer von Prä- bis Posttest zurückgeführt werden konnten. Im Modul 1 konnte zudem eine Wartegruppe realisiert werden, bei der, im Gegensatz zur „normalen“ Kontrollgruppe, von vergleichbaren Voraussetzungen (z. B. bezüglich der Motivation) wie in der Interventionsgruppe auszugehen war, sodass diesbezügliche Erklärungen nicht in Frage kamen.

Zusammengenommen ergeben die dargebrachten Informationen, dass die Schlussfolgerungen begründet vorgenommen und ihre Grundlagen transparent gemacht wurden. Der Standard kann daher als erfüllt gelten.

⇨ Standard G8 wurde erfüllt.

4.2.2.4.9 Meta-Evaluation (G9)

Um Meta-Evaluationen zu ermöglichen, sollen Evaluationen in geeigneter Form dokumentiert und archiviert werden.

Die DeGEval-Standards empfehlen, dass Evaluationen selbst evaluiert werden, also Meta-Evaluationen stattfinden. Denn misslungene Evaluationen können zu fehlerhaften Entscheidungen führen. Ebenso kann, womöglich zu Unrecht, starke Kritik an einer Evaluationsstudie geübt werden. Meta-Evaluationen können solche Situationen vermeiden helfen und haben den Zweck, die Qualität von Evaluationsprozessen und -ergebnissen zu bewerten und zu verbessern. Dabei können die Standards der DeGEval eingesetzt werden. Die Meta-Evaluation kann intern, also durch das Evaluationsteam, oder extern durch außenstehende Personen durchgeführt werden. Umfangreiche und vertiefte Meta-Evaluationen sind nur in bestimmten Fällen angezeigt, wohingegen knappe Selbstevaluationen durch die Evaluierenden in aller Regel stattfinden sollten. Wenn regelmäßig Meta-Evaluationen realisiert werden, erhöht dies die Glaubwürdigkeit einzelner Evaluationen sowie der Berufsgruppe der Evaluatoren. Damit Meta-Evaluationen, Evaluationssynthesen oder Meta-Analysen möglich sind, sind der Evaluationsgegenstand sowie die wesentlichen Zwecke, Schritte, Methoden, Daten und Ergebnisse von den Evaluatoren zu dokumentieren und zu archivieren. Das fördert den wissenschaftlichen Fortschritt und die Akkumulation von Wissen, sowohl im jeweiligen inhaltlichen Bereich als auch auf dem Gebiet der Evaluationsmodelle und -methoden. Weiterhin ist die wissenschaftliche Publikation über Evaluationsstudien und deren Methoden, Ergebnisse und Probleme wünschenswert, da sie einen wichtigen Beitrag zur Weiterentwicklung der Evaluation und zur Wissensverbreitung im jeweiligen Gegenstandsbereich leistet. Sie dient zudem der Qualitätsentwicklung sowie der Verbreitung und Akzeptanz von Evaluationen.

Im JCSEE-Handbuch wird ergänzend die formative von der summativen Meta-Evaluation unterschieden. Der formativen Meta-Evaluation kommt die Funktion zu, die Planung und Umsetzung einer Programmevaluation anzuleiten. Sie findet daher während der Programmevaluation statt und sollte idealerweise extern erfolgen, ist aber auch intern möglich. Die summative Funktion besteht in der Einschätzung der Qualität und Verwendbarkeit einer abgeschlossenen Programmevaluation. Dabei wird unter anderem der Frage nachgegangen, ob das Evaluationsdesign für die Studie angemessen war, oder ob die Schlussfolgerungen sich durch die Daten rechtfertigen lassen. Summative Meta-Evaluationen können ebenfalls extern erfolgen, alternativ kann ihre Durchführung durch zwei oder mehr Mitglieder des Evaluationsteams realisiert werden. Wünschenswert sind interne ebenso wie externe Sichtweisen sowie auch verschiedene externe Blickwinkel. Das Handbuch unterscheidet weiterhin formale von informalen Meta-Evaluationen, wobei formale Meta-Evaluationen durch eine systematische Vorgehensweise gekennzeichnet sind, während eine informale Meta-Evaluation z. B. lediglich anhand einer Checkliste erfolgen kann, die von den Evaluatoren regelmäßig ausgefüllt wird. Adressaten einer Meta-Evaluation sind alle Personen, deren Entscheidungsfindung mit der Evaluation zusammenhängt, sowie alle an der evaluierten Maßnahme Beteiligten und auch die Evaluatoren selbst. Schließlich gibt das Handbuch einige Empfehlungen die Rahmenbedingungen von Meta-Evaluationen betreffend. So sollten ausreichende finanzielle sowie sonstige Mittel für angemessene formative und summative Meta-Evaluationen eingeplant werden. Die Schlussredaktion des Meta-Evaluationsberichts sollte bei den Meta-Evaluatoren liegen. Die Adressaten dieses Berichts sollten ebenso wie seine Verbreitung bestimmt und aufgezeichnet werden. Da auch Meta-Evaluationen widerlegt oder evaluiert werden können, sollten die Aktivitäten, Informationen und Analysen dokumentiert werden.

Die Forderung nach wissenschaftlicher Publikation wurde bei der vorliegenden Evaluation durch den veröffentlichten Abschlussbericht erfüllt, zudem sind weitere Publikationen geplant. Evaluationsgegenstand, Zwecke, Schritte, Methoden, Daten und Ergebnisse wurden dokumentiert und sollen nach Möglichkeit auch archiviert werden, sodass die Grundlage für Meta-Evaluationen gegeben ist. Eine solche wurde auch bereits realisiert: Die vorliegende Arbeit befasst sich mit

der Evaluierung der Evaluation von „Elternchance ist Kinderchance“. Diese wurde durch ein Mitglied des Evaluationsteams, also intern durchgeführt. Es handelt sich um eine umfassende, summative Meta-Evaluation, die formal anhand der Standards für Evaluation der DeGEval (2008b), ergänzt durch das Handbuch zu den amerikanischen Evaluationsstandards (Yarbrough et al., 2011) erfolgte.

Da diese Meta-Evaluation im Rahmen einer Forschungsarbeit aus wissenschaftlichem Erkenntnisinteresse heraus durchgeführt wurde und es sich nicht um eine beauftragte Bewertung mit spezifischer inhaltlicher Zielsetzung handelte, sind einige der Forderungen dieses Standards nicht ganz zutreffend bzw. deren Nicht-Einhaltung kann mit diesem Umstand begründet und als angemessen bewertet werden. So wurde die vorliegende Meta-Evaluation als Qualifikationsarbeit realisiert, weshalb keine finanziellen oder sonstigen Mittel dazu notwendig waren; solche Mittel waren auch nicht veranschlagt gewesen. Diese Gegebenheit brachte es allerdings mit sich, dass die Bewertung nur intern durch die Autorin dieser Arbeit, die Mitglied des Evaluationsteams gewesen war, vorgenommen wurde, während eine Beurteilung durch externe Personen nicht erfolgte. Dem Standard zufolge sollen bei einer summativen Meta-Evaluation mindestens zwei Mitglieder des Evaluationsteams die Bewertung durchführen. In der vorliegenden Arbeit wurde so verfahren, dass die Bewertung durch die Meta-Evaluatorin zum einen durch Stellungnahmen verschiedener Stakeholder (darunter ein weiteres Mitglied des Evaluationsteams) ergänzt wurde. Zum anderen wurde die Bewertung der Meta-Evaluatorin diesem anderen Mitglied des Evaluationsteams sowie dem Auftraggeber vorgelegt und diese wurden um eine Einschätzung der Adäquatheit der Beurteilung gebeten (vgl. Abschnitt 4.2.1). Dieses Vorgehen konnte zwar nicht die Perspektive eines vom Bundesprogramm und seiner Evaluation unabhängigen Außenstehenden ersetzen, lieferte jedoch wichtige und wertvolle Informationen und reduzierte die Subjektivität der Bewertungen.

Weiterhin wurden keine Adressaten des Meta-Evaluationsberichts bestimmt und auch die Verbreitung dieses Berichts wurde nicht fest vereinbart. Da die Meta-Evaluation als Dissertation publiziert werden sollte, war der Zugriff für die zentralen Adressaten, nämlich die interessierte Forschungsgemeinde, gewährleistet. Auch eine Verbreitung an Beteiligte am Bundesprogramm und seiner Evaluation war vorgesehen, jedoch nicht schriftlich aufgezeichnet worden. Die Schlussredaktion des Berichts lag bei der Meta-Evaluatorin. Diese stellte mit ihrer

Arbeit eine detaillierte Beschreibung ihres Vorgehens, der verwendeten Informationen und ihrer Analyse zur Verfügung, sodass ihre Meta-Evaluation wiederum evaluiert werden könnte.

Zusammenfassend ist festzuhalten, dass bisher lediglich eine interne Meta-Evaluation stattfand, wobei für eine externe Überprüfung jedoch auch keine Notwendigkeit bestand. Die interne Meta-Evaluation diente dem wissenschaftlichen Erkenntnisgewinn und kann zudem als Ausgangspunkt für eine externe Evaluierung fungieren. Da auch die weiteren Anforderungen dieses Standards als eingehalten betrachtet werden können, ergibt sich folgende Bewertung:

⇨ Standard G9 wurde erfüllt.

4.2.3 Zusammenfassung der Ergebnisse der Meta-Evaluation

Die vorliegende Meta-Evaluation ergab, dass bei der Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ insgesamt 13 der 25 Evaluationsstandards eingehalten worden waren. Weitere sieben Standards wurden immerhin als *überwiegend erfüllt* und fünf Standards als *teilweise erfüllt* bewertet. Kein Standard erhielt die Bewertung *nicht erfüllt*. Daher lässt sich das Fazit ziehen, dass die Evaluation von „Elternchance ist Kinderchance“ eine gelungene Evaluationsstudie darstellt, die den Evaluationsstandards in erfreulichem Maße gerecht wird.

Eine Übersicht über die Bewertungen der 25 Einzelstandards enthält Abbildung 7. Auf die Bewertungsergebnisse hinsichtlich der vier Standardgruppen *Nützlichkeit*, *Durchführbarkeit*, *Fairness* und *Genauigkeit* wird im folgenden Diskussionsteil eingegangen. Dort werden auch die Stärken und Schwächen der Evaluation nochmals aufgezeigt und Empfehlungen daraus abgeleitet.

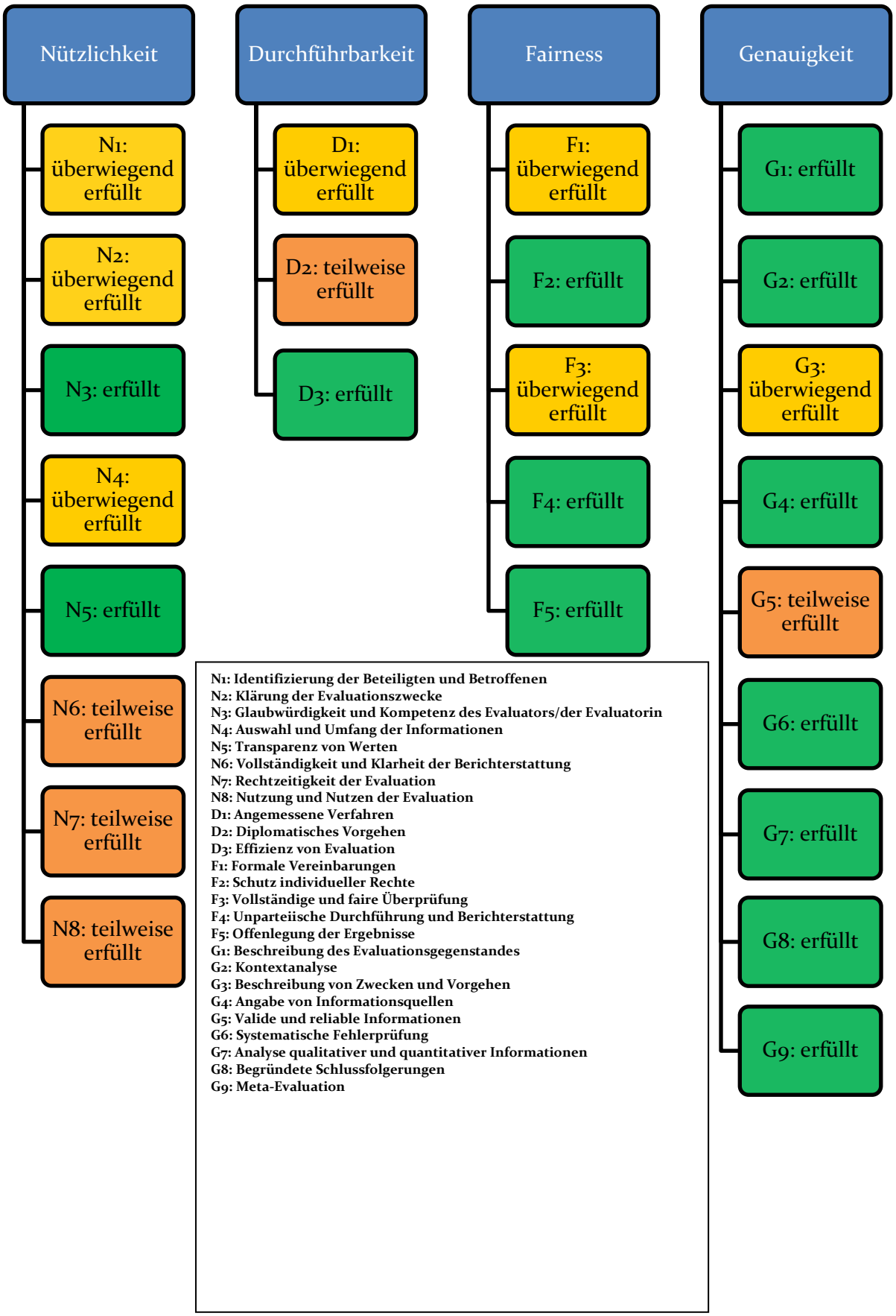


Abbildung 7: Übersicht über die Ergebnisse der Meta-Evaluation

5 Diskussion

In diesem Kapitel werden die Ergebnisse der Evaluationsbewertung diskutiert und in den theoretischen Hintergrund eingebettet. Zunächst erfolgt eine Darstellung der zentralen Herausforderungen im Evaluationsprozess, aus denen Empfehlungen für Evaluatoren und Auftraggebende von Evaluationen abgeleitet werden. Anschließend werden die Befunde der vorliegenden Studie im Hinblick auf die Evaluationsstandards diskutiert und auch hierzu Anregungen erarbeitet. Es folgt eine kritische Reflexion der angewendeten Methode, bevor im letzten Abschnitt ein Fazit gezogen und auf Forschungsimplikationen hingewiesen wird.

Zu beachten ist, dass die hier erarbeiteten Empfehlungen beabsichtigte Überschneidungen mit Anregungen aufweisen, die andere Autoren ausgesprochen haben. So stellen vor allem die in Abschnitt 5.1 gegebenen Empfehlungen vielfach eine Bestärkung von Hinweisen dar, die in den Evaluationsstandards bereits enthalten sind. Sie sind daher als Ergänzung bzw. Bekräftigung dieser Richtlinien zu verstehen. Da es sich bei der vorliegenden Arbeit um eine Fallstudie handelt, wurde außerdem die entsprechende Literatur gezielt herangezogen, um weitere Belege für die eigenen Ratschläge ausfindig zu machen und somit den Geltungsbereich zu erhöhen (Abschnitt 5.3.2). Schließlich sei darauf hingewiesen, dass dieses Kapitel Analogien zu Widmers Arbeit (1996) aufweist, sowohl inhaltlich (Aussprechen von Empfehlungen) als auch optisch (Hervorheben der Empfehlungen).

5.1 Zentrale Herausforderungen im Evaluationsprozess

In diesem Abschnitt werden die im Rahmen der Meta-Evaluation aufgedeckten zentralen Herausforderungen im Evaluationsprozess zusammengefasst und Empfehlungen dazu generiert. Die Darstellung folgt nicht der Reihenfolge der Evaluationsstandards, da einige Herausforderungen im Rahmen mehrerer Standards thematisiert wurden und Doppelungen nach Möglichkeit vermieden werden sollen. Aus Gründen der Übersichtlichkeit wurden stattdessen Kategorien erstellt, in denen die Herausforderungen nach inhaltlichen Gesichtspunkten

zusammengefasst geschildert werden. Die Typologie ist nicht als „trennscharf“ zu verstehen, da bei einigen Herausforderungen ein Bezug zu mehreren Kategorien hergestellt werden kann.

5.1.1 Besonderheiten des Evaluandums

Zunächst seien diejenigen Herausforderungen erwähnt, die in der Beschaffenheit des Evaluationsgegenstandes begründet lagen. Hierbei ist die heterogene Umsetzung der Qualifizierungsmaßnahme zu nennen, die auf die relativ offene Gestaltung der Curricula und das Ansetzen an den Ressourcen der Teilnehmenden zurückzuführen war. Dies stellt keinen Kritikpunkt am Bundesprogramm dar, sondern unter anderem diese Ressourcenorientierung war vielmehr eines der Alleinstellungsmerkmale dieses Programmes in der Familienbildungslandschaft. Diese Heterogenität schränkte allerdings die Evaluierbarkeit des Programmes ein (Standards N4 und G1).

Die Evaluierbarkeit war weiterhin dadurch beeinträchtigt, dass keine Explikation der vermuteten Wirkungsketten erfolgt war (Standard N4). Wie Stockmann ausführt, sollten Wirkungsevaluationen „auf einem theorie-basierten Konzept aufbauen, in dem klar die verschiedenen Interventionen, die erwarteten Wirkungen und die Rahmenbedingungen analysiert werden“ (Stockmann, 2010, S. 1). Daher wird folgende Empfehlung für Auftraggeber von Evaluationsprojekten ausgesprochen:

> **Empfehlung 1**: Auftraggebende von Evaluationen sollten sich vor der Ausschreibung von den Programm- bzw. Maßnahmenverantwortlichen die Wirkungsketten aufzeigen lassen und den Evaluatoren das Wirkmodell darlegen.

Des Weiteren wurde im Rahmen der Meta-Evaluation deutlich, dass die verschiedenen Stakeholder unterschiedliche Erwartungen an die Evaluation hatten, was im Einklang mit den im Theorieteil angeführten Erläuterungen zu impliziten Nutzenerwartungen steht (Abschnitt 2.1.6.1). Der im Rahmen der Standards N5 und D2 angesprochene Befund, wonach ein Qualifizierungsträger eine vergleichende Auswertung der beiden Curricula präferiert hätte, der andere hingegen nicht, zeigt, dass bei einer Evaluation nicht immer den Interessen aller Stakeholder

entsprochen werden kann, da diese sogar innerhalb einer Stakeholdergruppe divergieren können (vgl. Brandt, 2007). Bei der vorliegenden Evaluation hätten die Erwartungen der Träger womöglich zu Beginn der Evaluation noch direkter angesprochen und es hätte nachfolgend mit dem Auftraggeber geklärt werden können, inwieweit die Evaluation diese Erwartungen erfüllen sollte.

> **Empfehlung 2**: Evaluatoren sollten zu Beginn der Evaluation die Erwartungen der Stakeholder in Erfahrung bringen. Sie sollten mit dem Auftraggeber klären, inwieweit diesen Erwartungen entsprochen werden kann und soll, und dies den Stakeholdern offen kommunizieren.

Zu beachten ist hierbei, dass die Kooperationsbereitschaft der Stakeholder sinken könnte, wenn ihre Erwartungen an die Evaluation nicht erfüllt werden sollen. In solchen Fällen sollte durch den Auftraggeber, die Evaluatoren und die betreffenden Stakeholder ein sinnvoller Kompromiss erzielt werden.

Schließlich seien die sehr mäßigen Rücklaufquoten der Kontrollgruppe genannt, die ebenfalls eine große Herausforderung dieses Projekts (und vermutlich auch anderer Wirkungsevaluationen) darstellten und in deutlich erhöhtem Arbeitsaufwand für das Evaluationsteam resultierten (Standard N7). Da die Bemühungen der Evaluatoren zur Behebung dieser Probleme wenig erfolgreich waren, werden hieraus keine Empfehlungen abgeleitet. Es wird allerdings angeregt, für solche oder ähnliche Schwierigkeiten mehr Pufferzeiten einzuplanen, als dies im vorliegenden Projekt der Fall war (vgl. hierzu den folgenden Abschnitt).

5.1.2 Planung und Durchführung der Evaluation

In diesem Abschnitt werden Ratschläge für die Planung und Durchführung einer Evaluation erarbeitet. Auch auf vertragliche Vereinbarungen wird eingegangen.

Zunächst werden die Verzögerungen aufgegriffen, die sich im Projektverlauf ergaben und die zu beträchtlichem Mehraufwand für das Evaluationsteam führten, sodass insgesamt die geplanten Ressourcen

als zu gering für die Bewältigung der Evaluationsaufgaben zu bewerten waren (Standards N7 und D1). Hinzu kam ein etwas zu ambitioniertes Evaluationsdesign mit zahlreichen aufwendigen Datenerhebungen (Standard N4). Wenngleich die Verzögerungen nicht vorhersehbar waren und daher auch nicht eingeplant werden konnten, könnte der hier zu verzeichnende Ressourcenmangel in künftigen Evaluationen durch eine realistische Anlage der Evaluation sowie durch das Berücksichtigen eines ausreichenden zeitlichen Puffers vermieden werden.

Empfehlung 3: Evaluationen sollen realistisch geplant werden. Insbesondere sind im Zeitplan der Evaluierenden unbedingt Pufferzeiten für mögliche Schwierigkeiten und Verzögerungen vorzusehen.

Zur realistischen Zeitplanung gehört weiterhin, dass die Evaluation auf die Lieferung von Ergebnissen zu wichtigen Entscheidungen hin ausgerichtet sein soll. In der vorliegenden Evaluation waren diesbezüglich gewisse Abstriche zu verzeichnen (Standard N7), was zu folgender Empfehlung veranlasst:

Empfehlung 4: Evaluationen sollen ihre Ergebnisse rechtzeitig zu anstehenden Entscheidungen erbringen. Um dies zu ermöglichen, sollen Auftraggebende Entscheidungen absehen und den Evaluatoren kommunizieren, damit diese die Evaluationsaktivitäten darauf ausrichten können. Dies gilt auch für Entscheidungen, die sich erst im Evaluationsverlauf ergeben. Den Evaluierenden kommt die Aufgabe zu, den Auftraggebern die Relevanz einer rechtzeitigen Evaluation zu verdeutlichen.

Im Rahmen der Planung und Durchführung kamen zwei Verfahren zum Einsatz, die vom Handbuch des JCSEE empfohlen werden: Es wurde eine Pilotstudie durchgeführt und es fand eine Hospitation bei einem Qualifizierungskurs statt. Diese beiden Aktivitäten erwiesen sich als sehr hilfreich. Durch die Hospitation konnten wichtige Eindrücke von der Maßnahme gewonnen werden (Standard G1) und die Pilotstudie ermöglichte den Einbezug von Elternbegleitern und Eltern in die Planung der Erhebungsinstrumente sowie eine Prüfung derselben auf Verständlichkeit und Durchführbarkeit (Standards N1, N3, N4,

D1 und G6). Daher werden die Empfehlungen des Handbuchs bekräftigt:

> **Empfehlung 5**: Evaluierende sollten eine Pilotstudie durchführen. Diese ermöglicht neben der Testung der Erhebungsinstrumente auf Durchführbarkeit und Verständlichkeit den Einbezug der Gruppe der zu Befragenden.

> **Empfehlung 6**: Sofern möglich, sollten Evaluatoren zu Beginn der Evaluation die zu evaluierende Maßnahme beobachten, um einen genaueren Eindruck davon zu erhalten.

Einen Aspekt, der diskutiert werden soll, auch wenn dazu keine Empfehlungen generiert werden, stellt die „unpersönliche" Rekrutierungsstrategie der vorliegenden Evaluation dar. Die Fachkräfte wurden per E-Mail kontaktiert, die Eltern wurden von Fachkräften um Teilnahme gebeten und an das Evaluationsteam weitergeleitet (Standard N3). Es gab somit keinen persönlichen Kontakt in dem Sinne, dass das Evaluationsteam in einer Einrichtung vor Ort das Projekt vorgestellt und bei Fachkräften und Eltern für die Teilnahme an der Befragung geworben hätte. Nun ist einerseits denkbar, dass die erwähnte geringe Rücklaufquote, insbesondere in der Kontrollgruppe, auf diesen „unpersönlichen" Rekrutierungsweg zurückzuführen sein könnte und durch eine persönlichere Strategie hätte erhöht werden können (Standard D1). Andererseits ist zu bedenken, dass externe Evaluierende nicht grundsätzlich auf Akzeptanz stoßen und sogar Abwehrreaktionen hervorrufen können, da sie von außen kommen (Abschnitt 2.1.3.4). In diesem Fall wäre die Ansprache durch die vertraute Fachkraft der geeignetere Weg, zumindest für die Elternstichprobe.

Die Durchführung der Befragungen, die bei den Fachkräften in Form von Online-Fragebögen und bei den Eltern mittels telefonischer Interviews stattgefunden hatten, erwies sich als sehr gut zu den Bedarfen der Teilnehmenden passend und wurde von den Stellungnehmern sehr positiv bewertet (Standard D1). Da die angemessenen Befragungsformen jedoch von den Rahmenbedingungen der zu Befragenden abhängen, welche bei jeder Evaluation unterschiedlich sein dürften, wird daraus keine Empfehlung abgeleitet. Die passende Form der Befragung

kann jedoch im Rahmen einer Pilotstudie ermittelt werden (Empfehlung 5).

Es gab eine einzige negative Rückmeldung zur Elternbefragung, die jedoch in Relation zu den vielen positiven Angaben zu sehen ist (Standard D1). Aufschlussreich ist hierbei, dass die betreffende Person diese Rückmeldung nicht direkt ihrem Interviewer gegenüber ausdrückte, sondern diese die Evaluatoren erst über die Fachkraft und einen Kursleiter vermittelt erreichte, obwohl nach jedem Interview nachgefragt wurde, ob der Befragte noch Anmerkungen jedweder Art habe. Wenngleich nicht damit zu rechnen ist, dass auf eine solche Frage offene Kritik dem Interviewer gegenüber geäußert wird (wie auch dieser Fall belegt), erscheint die Frage nach Anmerkungen des Befragten zum Ende des Interviews dennoch sinnvoll.

Empfehlung 7: Am Ende einer Befragung sollte den Befragten die Möglichkeit gegeben werden, Fragen, Anmerkungen oder auch Kritik zu äußern.

Die multidisziplinäre Zusammensetzung des Evaluationsteams erwies sich in allen Phasen des Evaluationsprozesses als sehr gewinnbringend, da die unterschiedlichen Hintergründe und Perspektiven zu wertvollen Diskussionen beitrugen und einseitigen Bewertungen oder Interpretationen entgegenwirkten (Standard N3). Daher wird folgende Empfehlung gegeben:

Empfehlung 8: Evaluationsteams sollten sich, wenn möglich, aus Personen mit unterschiedlichen Ausbildungs- und Erfahrungshintergründen zusammensetzen.

Schließlich soll auf die vertraglichen Vereinbarungen eingegangen werden. Bei der vorliegenden Evaluation waren sehr viele der in den Evaluationsstandards empfohlenen Punkte vertraglich geregelt (Standard F1). Die Ausnahmen, also die nicht schriftlich festgelegten Gesichtspunkte, die jedoch von den Standards genannt werden, werden im Folgenden diskutiert.

In den Standards wird empfohlen, die Verantwortlichkeiten des Personals der zu evaluierenden Einrichtung bei der Durchführung der Evaluation sowie bei der Teilnahme an der Evaluation vertraglich zu

regeln. Des Weiteren sollen auch Dienstleistungen, Personal, Informationen und Materialien, die vom Auftraggeber bereitgestellt werden (inklusive Datenzugang), schriftlich festgelegt werden. Wie beim Standard F1 erläutert, waren diese beiden Bestimmungen für die vorliegende Evaluation nicht ganz passend, da Verträge nur zwischen Auftraggeber und Evaluatoren bzw. zwischen den beiden Parteien des Evaluationsteams geschlossen worden waren. Die Datenerhebung fand jedoch bei Fachkräften und Eltern in ganz Deutschland statt und der Datenzugang wurde nicht vom Auftraggeber vermittelt. Wenn es sich um eine Erhebung in bestimmten Einrichtungen gehandelt hätte, hätten ggf. mit diesen schriftliche Vereinbarungen getroffen werden können. Es bleibt festzuhalten, dass nicht jede Anforderung der Standards bei jeder Evaluation zutreffend ist, worauf auch in den Standards selbst hingewiesen wird (Abschnitt 0) und was sich hier ebenfalls zeigte.

Weitere drei Aspekte waren nicht, wie von den Standards nahegelegt, schriftlich festgehalten, aber dennoch im Rahmen der Evaluation eingehalten worden. Dies waren die Ermittlung der Beteiligten und Betroffenen und die Methoden der Qualitätskontrolle bei der Datenerhebung, Auswertung und Berichterstattung (Standard F1) sowie die Offenlegung der Ergebnisse (Standard F5). Dennoch wurden die Beteiligten und Betroffenen erfolgreich ermittelt (Standard N1), es wurde sehr auf Qualität geachtet und diese durch viele Maßnahmen sichergestellt (Standard F1) und die Ergebnisse wurden durch die Veröffentlichung des Abschlussberichts offengelegt (Standard F5). Dass die Einhaltung auch nicht vertraglich festgelegter Gesichtspunkte im vorliegenden Projekt gut funktionierte, bedeutet jedoch nicht, dass dies auch bei anderen Evaluationen der Fall sein wird. Es wird daher geraten, alle von den Standards empfohlenen Aspekte schriftlich zu vereinbaren, sofern sie für eine Evaluation passend sind.

Empfehlung 9: Die von den Evaluationsstandards empfohlenen Vertragsinhalte sollten von Auftraggebenden und Evaluatoren auf Passung für das Evaluationsprojekt geprüft und *alle voraussichtlich zutreffenden* Gesichtspunkte sollten schriftlich vereinbart werden.

Weiterhin waren weder die Festlegung der Adressaten der Berichte und der Darstellungsmethoden, noch die Gegendarstellung durch die

Evaluierten ein Bestandteil der vertraglichen Vereinbarungen (Standard F1). Zudem war hinsichtlich der Ergebnisverbreitung zwar das Recht, aber nicht die Pflicht festgehalten (Standards N7 und D2). Da bei der vorliegenden Evaluation gewisse Mängel bei der Ergebnisverbreitung (Standard N8) sowie bei der Passung der Berichtsformen für die Stakeholdergruppen (Standard N6) zu verzeichnen waren, ist das Fehlen der genannten Punkte kritisch zu bewerten. So hätte eine schriftliche Übereinkunft zu den Adressaten der Berichte und den Darstellungsmethoden sowie zur Verbreitung der Ergebnisse womöglich zu einer besseren Berücksichtigung der Berichtsbedarfe der verschiedenen Stakeholder geführt. Eine entsprechende Empfehlung wurde bereits ausgesprochen (Empfehlung 9). Allerdings ist hier die Auftraggeberorientierung der Evaluation zu bedenken, aufgrund welcher eine Gegendarstellung durch die Evaluierten und die Ergebnisverbreitung nicht näher vereinbart wurden. Die Herausforderungen, die sich durch den auftraggeberorientierten Ansatz für den angemessenen Einbezug der Beteiligten ergaben, werden im folgenden Abschnitt diskutiert.

5.1.3 Berücksichtigung der Stakeholder

In den DeGEval-Standards wird darauf hingewiesen, dass nicht alle Standards bei jeder Evaluation eingehalten werden können, unter anderem aufgrund der vielfältigen Rahmenbedingungen und Praxisanforderungen (Abschnitt 0). Bei der vorliegenden Meta-Evaluation wurden insbesondere zwei Rahmenbedingungen deutlich, an denen die (Nicht-)Erfüllung von Standards mitunter relativiert werden musste. Dies waren zum einen die zur Verfügung stehenden Ressourcen (z. B. Standards N1, D1, G4) und zum anderen die Tatsache, dass ein auftraggeberorientierter Evaluationsansatz verfolgt wurde (z. B. Standards N4, N5, D2, F4). Wie im Methodenteil (Abschnitt 4.2.1) erläutert wurde, hatte die Orientierung an den Evaluationsstandards zunächst zu einer etwas strengeren Bewertung einiger Standards geführt, weil die Evaluationsstandards die Beteiligung der Stakeholder sehr betonen. Auf die Frage, inwieweit die Evaluationsstandards für verschiedene Evaluationsansätze gleichermaßen Gültigkeit besitzen, wird unten eingegangen (Abschnitt 5.2.1). In diesem Abschnitt werden die Herausforderungen aufgezeigt, die sich aus dem Spannungsfeld zwischen

auftraggeberorientierter Evaluation und dem angemessenen Einbezug der Beteiligten ergaben.

5.1.3.1 Beteiligung der Stakeholder bei der Evaluationsplanung und -durchführung

Hierbei ist zunächst die angemessene Beteiligung der Stakeholder bei der Planung und Durchführung der Evaluation zu nennen (Standards N1 und N4). Diese wurde als überwiegend gelungen bewertet in Anbetracht der Tatsache, dass ein auftraggeberzentriertes Vorgehen verfolgt wurde (die Evaluationsstandards bieten allerdings keine konkreten Anhaltspunkte dafür, welches Ausmaß an Beteiligung bei einer auftraggeberorientierten Evaluation angemessen ist; vgl. hierzu Abschnitt 5.2.1). Wenngleich alle Stellungnehmer ihre Beteiligung am Evaluationsprozess als gut oder sogar sehr gut einschätzten, gab es doch einige Verbesserungsvorschläge, aus denen auch Empfehlungen abgeleitet werden können.

So hatten der Vertreter eines Trägers und der Kursleiter dieses Trägers sich eine Rückmeldung zu Stärken und Schwächen ihrer Qualifizierung erhofft, jedoch nicht erhalten. Aus Sicht des Trägers hätte dies mit einer besseren Beteiligung an der Planung der Erhebungsinstrumente womöglich umgangen werden können. Zudem sei auf Empfehlung 2 verwiesen, denn bei frühzeitiger und gezielter Klärung der Erwartungen wären diese bekannt gewesen und dem Träger hätte kommuniziert werden können, inwieweit seinen Erwartungen entsprochen werden kann.

Der Vorschlag der Servicestelle, zu Beginn der Evaluation mit allen Stakeholdern ein Informationstreffen im Hinblick auf die Planung und Durchführung abzuhalten, erscheint sehr sinnvoll, um den Austausch zwischen Evaluierenden und Stakeholdern, sowie auch zwischen den Stakeholdergruppen zu fördern. Durch ein solches Treffen wäre vermutlich auch deutlich geworden, dass den Dozenten nähere Informationen zur Verfügung hätten gestellt werden müssen, damit diese bei ihren Teilnehmern besser für die Befragung hätten werben können (wie der Vertreter eines Trägers in seiner Stellungnahme anregte). Es wird daher Folgendes empfohlen:

Empfehlung 10: Zu Beginn einer Evaluation sollte ein Informationstreffen bezüglich Planung und Durchführung mit den Stakeholdern stattfinden.

Ein Trägervertreter äußerte zudem, dass im Nachhinein betrachtet mehr Austausch mit der FAU im Evaluationsverlauf wünschenswert gewesen wäre. Mit dem DJI hatte unter anderem aufgrund der Datenerhebung mehr Kooperation bestanden. Für den Fall, dass die Evaluation negative Ergebnisse erbracht hätte, hätte man sich womöglich auf den zu geringen Austausch berufen. Daher wird im Sinne der Beteiligung der Stakeholder sowie der Transparenz folgender Ratschlag gegeben:

Empfehlung 11: Sind die Zuständigkeiten im Evaluationsteam aufgeteilt, sollte dennoch von allen Gruppen des Evaluationsteams der Austausch mit den relevanten Stakeholdern gepflegt werden.

Hinsichtlich der methodischen Konstruktion der Erhebungsinstrumente ist festzuhalten, dass die Kompetenzen hierfür beim Evaluationsteam liegen. In inhaltlicher Hinsicht hätten jedoch Elternbegleiter bei der Erstellung des Elternfragebogens einbezogen werden können, um die Perspektive der Praxis einzufangen. Dies wurde auch von einigen der befragten Stellungnehmer befürwortet. Wenngleich im Rahmen der Stellungnahmen lediglich die Beteiligung von Elternbegleitern bei der Konstruktion des Elternfragebogens angesprochen wurde, erscheint die Beteiligung von Dozenten bei der Konstruktion des Elternbegleiterfragebogens ebenso sinnvoll. Neben den Rückmeldungen der Elternbegleiter und Eltern, die sie im Rahmen der Pilotstudie zu ihren eigenen Befragungen gegeben hatten (gleiche Ebene), hätten somit auch Vertreter der jeweils „höher gelegenen Ebene“ mit einbezogen werden können.

Empfehlung 12: Bei der Konstruktion von Erhebungsinstrumenten sollten Rückmeldungen von Vertretern der Praxis eingeholt werden. Liegen mehrere Erhebungsebenen vor, können die Rückmeldungen auf gleicher Ebene, die im Rahmen einer Pilotstudie gegeben werden, durch Rückmeldungen der übergeordneten Ebene ergänzt werden.

5.1.3.2 Passende Berichtsformen

Dem Standard N6 zufolge sollen Evaluationsberichte vollständig und leicht zu verstehen sein. Zudem ist für jede Zielgruppe eine angemessene Vermittlungsform zu wählen. Wie die Stellungnahmen im Rahmen der Meta-Evaluation zeigten, passten die Berichtsformen der vorliegenden Evaluation nur teilweise zu den Bedürfnissen der verschiedenen Beteiligten. Auch hier fehlen Hinweise der Evaluationsstandards, inwieweit neben dem Auftraggeber als zentralem Nutzer der Evaluation die Bedürfnisse anderer Stakeholder berücksichtigt werden dürfen oder müssen. Es erscheint sinnvoll, die verschiedenen Beteiligten zu Beginn der Evaluation auf ihre Informationsbedürfnisse sowie die gewünschten Berichtsformen anzusprechen, die Machbarkeit dieser Wünsche mit dem Auftraggeber zu klären und den hierzu gefassten Beschluss den jeweiligen Stakeholdern mitzuteilen.

> **Empfehlung 13**: Zu Beginn der Evaluation sollten die Evaluatoren die Informationsbedürfnisse und die gewünschten Berichtsformen der verschiedenen Beteiligten in Erfahrung bringen, diese Wünsche hinsichtlich ihrer Umsetzbarkeit mit dem Auftraggeber besprechen und den Beteiligten die Entscheidung rückmelden.

Hinsichtlich der vollständigen und klaren Berichterstattung ist zu bemerken, dass laut JCSEE-Handbuch die Bemühungen, die zugunsten der Klarheit unternommen werden, nicht die Genauigkeit der Evaluation beeinträchtigen dürfen. Es erscheint widersprüchlich, dass ein Bericht einerseits vollständig sein und wissenschaftlichen Standards entsprechen und andererseits leicht verständlich aufbereitet werden soll. Dies zeigte sich im vorliegenden Projekt z. B. darin, dass es sich als herausfordernd erwies, die Ergebnisse komplexer Analysen in einfachen Worten darzustellen, die dennoch das Ergebnis korrekt wiedergaben (Standard G7). Hierin spiegelt sich das Spannungsverhältnis zwischen den Anforderungen der Wissenschaft und denen der Praxis wider, welches im Theorieteil besprochen wurde (Abschnitt 2.1.7.1). Es wird in Anlehnung an Erzberger und Kelle (2014) empfohlen, dass Evaluatoren sich als angewandte Forscher verstehen und, um ihrem Auftrag einer nützlichen Studie nachzukommen, ihre wissenschaftlich fundierten Ergebnisse verständlich vermitteln.

Empfehlung 14: Evaluatoren sollten sich ihrer Rolle als angewandte Auftragsforscher bewusst sein. Das bedeutet, dass sie nicht nur wissenschaftlich belastbare Ergebnisse erzielen, sondern auch für verständliche Kommunikation sorgen müssen, da nur verstandene Ergebnisse genutzt werden können.

In der vorliegenden Evaluation wurde für den Widerspruch zwischen Klarheit und Genauigkeit eine praktikable Lösung gefunden. Die beschriebene Schwierigkeit wurde dadurch umgangen, dass mit der Kurzbroschüre ein leicht verständliches Medium erstellt wurde, während der Abschlussbericht den Anspruch der Genauigkeit erfüllte (Standard N6). Daher wird, angelehnt an Widmer (1996), die Erstellung mehrerer Evaluationsberichte empfohlen, sofern die verfügbaren Ressourcen dies zulassen.

Empfehlung 15: Um dem Anspruch eines sowohl genauen als auch verständlichen Berichts gerecht zu werden, können Evaluatoren zwei Berichte verfassen, von denen einer auf hohe Verständlichkeit abzielt und der andere das Kriterium der Genauigkeit erfüllt.

Erwähnenswert ist hierbei, dass die Erstellung der Kurzbroschüre eine Idee des Auftraggebers war. Dies verdeutlicht, dass nicht nur die Evaluatoren für die Einhaltung der Evaluationsstandards verantwortlich sind, sondern auch andere Stakeholder dazu beitragen können. Dieser Gedanke wird in Abschnitt 5.2 aufgegriffen.

5.1.3.3 Information über Ergebnisse

Im Hinblick auf (Zwischen-)Ergebnisse der Evaluation wurde deutlich, dass diese zum Zeitpunkt der Stellungnahmen nur einem Teil der Stakeholder bekannt waren (Standard N7). Angesichts der Tatsache, dass die Kurzbroschüre schon einige Zeit veröffentlicht gewesen war, ist dies durchaus als kritisch zu sehen. Besonders die Äußerung eines Kursleiters, er selbst habe relativ viel Kontakt zu den zuständigen Mitarbeitern und dennoch keine Informationen erhalten, weist darauf hin, dass die Kommunikation der Ergebnisse nicht wie von den Evaluatoren und dem Auftraggeber angenommen funktionierte. Diese waren davon ausgegangen, dass die Informationen weitergeleitet würden, z. B. von

den Leitungen der Träger an die Kursleiter. Daher waren vom Evaluationsteam auch nur mit den Leitungen der beiden Träger Workshops organisiert worden, nicht hingegen mit den jeweiligen Dozenten. Die Evaluatoren hätten durch Nachfragen oder durch entsprechende Übereinkünfte sicherstellen müssen, dass die Informationen die intendierten Nutzer auch erreichten. Alternativ hätten sie die Informationsvermittlung an die betreffenden Stakeholdergruppen selbst übernehmen müssen.

> **Empfehlung 16**: Die Evaluatoren dürfen nicht davon ausgehen, dass Ergebnisse und andere Informationen weitergeleitet werden, sondern müssen deren Vermittlung sicherstellen oder selbst übernehmen.

Auch in Bezug auf die Ergebnisverbreitung bleibt seitens der Evaluationsstandards unklar, wie diesbezüglich die Ressourcen aufzuteilen sind, wenn der Auftraggeber den zentralen Nutzer darstellt, andere Stakeholder aber ebenfalls über Ergebnisse informiert werden wollen.

5.1.3.4 Unterstützung bei der Ergebnisnutzung

Hinsichtlich der Unterstützung bei der Nutzung der Evaluationsergebnisse empfehlen die Standards, Nachbereitungsveranstaltungen mit den Stakeholdern abzuhalten (Standards N6 und N8). Diesem Ratschlag konnte in der vorliegenden Evaluation nicht nachgekommen werden, weil für solche Workshops keine Ressourcen eingeplant waren. Das Ende der Evaluation und damit der Verfügbarkeit der Evaluierenden fiel mit der Abgabe des Abschlussberichts zusammen, und es waren weder vor noch nach der Abgabe solche Veranstaltungen eingeplant. Auch hier lassen die Standards offen, wie sehr die Nutzung durch verschiedene Stakeholder unterstützt werden soll, wenn der zentrale Nutzer der Auftraggebende ist. Jedoch dürfte der Nutzen einer Evaluation mit zunehmender Nutzung der Ergebnisse durch verschiedene Beteiligte steigen, was ein wünschenswertes Ziel darstellt. Da zudem von mehreren Stellungnehmern der Wunsch nach einer nachbereitenden Veranstaltung geäußert wurde, wird hier die Empfehlung der Standards bekräftigt:

Empfehlung 17: Um die Nutzung der Ergebnisse zu unterstützen, sollten Nachbereitungsveranstaltungen mit den Stakeholdern stattfinden. Die dafür erforderlichen zeitlichen und finanziellen Ressourcen sind von Auftraggeber und Evaluierenden einzuplanen. Die Evaluatoren sollten den Auftraggebenden die Notwendigkeit solcher Veranstaltungen nahebringen.

5.1.3.5 Zusammenfassende Hinweise

In den vorigen vier Abschnitten wurden einige Spannungen zwischen der Berücksichtigung des Auftraggebers als zentralem Evaluationsnutzer und der angemessenen Beteiligung der weiteren Stakeholder deutlich. Die Evaluationsstandards bieten diesbezüglich leider kaum Orientierungshilfen für (Meta-)Evaluatoren (vgl. hierzu auch Abschnitt 5.2.1). Damit bleibt unklar, welche Pflichten Evaluierende bei auftraggeberorientierten Evaluationen den anderen Stakeholdern gegenüber haben. Zu beachten ist, dass es für die beschriebene Problematik wohl keine Patentlösung geben kann und eine solche auch von den Evaluationsstandards nicht zu erwarten ist. Vielmehr soll hiermit eine Diskussion zu dieser Fragestellung angeregt werden, denn es wäre von Interesse, wie in vergleichbaren Studien mit dieser Thematik umgegangen wird.

5.1.4 Glaubwürdigkeit und Diplomatie

Die Evaluatoren des vorliegenden Projekts waren von den Beteiligten in sehr hohem Maße als glaubwürdig und kompetent wahrgenommen worden (Standard N3). Da die zu diesem Zweck durchgeführten Maßnahmen somit offenbar erfolgreich gewesen waren, lassen sich daraus einige Empfehlungen ableiten. Neben dem regelmäßigen Austausch mit dem Auftraggeber scheinen auch die Arbeitstreffen mit den beiden Qualifizierungsträgern, bei denen unter anderem aktuelle Zwischenergebnisse vorgestellt wurden (was für die Evaluatoren einigen Aufwand bedeutete), viel zur Glaubwürdigkeit des Evaluationsteams beigetragen zu haben. Auch die Teilnahme der Evaluatoren an Veranstaltungen der Träger zeugte von Engagement und der Relevanz, die der Austausch mit den Trägern für sie hatte.

Empfehlung 18: Im Sinne der Glaubwürdigkeit sollten Evaluatoren regelmäßigen Austausch mit dem Auftraggeber und weiteren Stakeholdern pflegen.

Des Weiteren hatten die Evaluierenden den Befragten ihre Ansprechbarkeit signalisiert, indem sie diesen ihre Kontaktdaten vermittelten. Einen deutlich höheren Beitrag zu ihrer Glaubwürdigkeit dürfte jedoch das datenschutzkonforme Vorgehen der Evaluatoren geleistet haben, für die die vertrauliche Behandlung von Daten und die Rechte der Teilnehmenden höchste Priorität hatten (Standard F2). Dieses Vorgehen entsprach rechtlichen und ethischen Grundsätzen, sodass sich folgende Empfehlung für künftige Evaluationen ergibt:

Empfehlung 19: Evaluierende sollten dem Datenschutz und den Rechten der Befragten oberste Priorität einräumen, um glaubwürdig zu agieren und rechtlichen sowie ethischen Grundsätzen zu entsprechen.

Unter dem Aspekt der Glaubwürdigkeit kann auch Empfehlung 10 aufgegriffen werden, wonach am Anfang eines Evaluationsprojekts ein informierendes Treffen mit den Stakeholdern stattfinden sollte. Wenn dort die Evaluationsplanung offen kommuniziert wird und die Evaluierenden als Ansprechpartner für Fragen zur Verfügung stehen, erhöht dies sicherlich die Akzeptanz durch die Stakeholder.

Im Sinne der Glaubwürdigkeit sowie auch des diplomatischen Verhaltens sorgten die Evaluatoren dafür, dass mit beiden Qualifizierungsträgern gleich viel Austausch bestand und beide Parteien die gleichen Informationen erhielten (Standard D2). Dies war deshalb besonders wichtig, weil zwischen den Trägern ein Konkurrenzverhältnis bestand und das Evaluationsteam keinen Zweifel an seiner Neutralität aufkommen lassen wollte, indem etwa ein Träger bevorzugt behandelt worden wäre. Tatsächlich bestand mit den Trägern eine sehr gute Kooperation und im Rahmen der gemeinsamen Treffen konnten beiläufig einige Bedenken herausgefunden werden. Allerdings hätten die Evaluatoren mögliche Bedenken direkter erfragen können, wozu auch im JCSEE-Handbuch geraten wird.

Empfehlung 20: Wird zwischen Stakeholdern Konkurrenz vermutet, so sollten sich die Evaluierenden mit den betreffenden Stakeholdern in gleichem Maße austauschen, um Glaubwürdigkeit und Unparteilichkeit zu signalisieren.

Empfehlung 21: Die Evaluatoren sollten sich bei den Stakeholdern direkt nach möglichen Bedenken erkundigen.

Bezogen auf den Aspekt der Diplomatie empfehlen die DeGEval-Standards weiterhin, dass die Evaluierenden den Dialog zwischen Entscheidern und anderen Stakeholdern unterstützen sollen (Standard D2). Im Rahmen der Meta-Evaluation wurde deutlich, dass dieser Ratschlag in der vorliegenden Evaluation etwas besser hätte befolgt werden können (Standards N2, D2 und G3), sodass diese Empfehlung bekräftigt werden kann:

Empfehlung 22: Evaluatoren sollen es als ihre Aufgabe verstehen, die Kommunikation zwischen dem Auftraggeber und anderen Stakeholdern zu fördern.

Dem Handbuch des JCSEE zufolge sollen Evaluierende sich möglicher Einflussnahmen durch Beteiligte bewusst sein, um ggf. darauf reagieren zu können (Standard D2). Diese Anregung soll ergänzt werden, indem empfohlen wird, dass Evaluationsteams bewusst reflektieren, ob Einflussnahmen wahrgenommen werden. Dadurch könnten Beeinflussungsversuche noch besser aufgedeckt werden und es dürfte ihnen gezielter begegnet werden können. Bei Evaluationen, die nur von einer Person durchgeführt werden, könnte stattdessen die Meinung von Kollegen (im Sinne einer Supervision) eingeholt werden.

Empfehlung 23: In Evaluationsteams sollte gezielt reflektiert werden, ob Einflussnahmen von jeglicher Seite wahrgenommen wurden. Bei Evaluationen, die von Einzelpersonen durchgeführt werden, kann ersatzweise der Rat von Kollegen eingeholt werden.

Auch in der vorliegenden Evaluation wurden Beeinflussungsversuche durch Stakeholder wahrgenommen, die sich auf deren jeweilige Interessen zurückführen ließen, sodass sie in gewisser Weise

abzusehen waren (Standards D2 und F4). Während Evaluatoren sich einerseits nicht einseitig beeinflussen und instrumentalisieren lassen dürfen, sollten sie andererseits unverhältnismäßige von nachvollziehbarer bzw. sinnvoller Einflussnahme unterscheiden. Denn die detaillierten Korrekturen, die der Auftraggeber an den Berichten vornahm, stimmten zwar zunächst skeptisch, erwiesen sich aber weitgehend als hilfreich und nur teilweise konnten Änderungsvorschläge nicht umgesetzt werden, weil sie sich mit den Daten nicht vereinbaren ließen. Zu beachten ist, dass der Auftraggeber sich in der Regel besser mit dem Evaluandum und den Gegebenheiten in dessen Umfeld auskennt als externe Evaluatoren und deshalb besser einschätzen kann, welche Formulierungen womöglich falsch verstanden werden könnten (Beywl, 2012). Daher sollten Evaluierende zwar nach wie vor wachsam gegenüber möglichen Einflussnahmen sein und Korrekturen an Berichten nicht unkritisch hinnehmen. Ihnen sollte jedoch auch bewusst sein, dass, um es mit Beywls Worten auszudrücken, „nicht jeder kritische Einwand von Auftraggebenden … ein unstatthafter Beeinflussungsversuch [ist]“ (Beywl, 2012, S. 16).

5.1.5 Genauigkeit vs. Durchführbarkeit

Abschließend sei auf die für die vorliegende Evaluation zu beobachtenden Widersprüche zwischen verschiedenen Standards bzw. Standardgruppen eingegangen. In der Literatur werden solche Konflikte z. B. für die Genauigkeitsstandards und die Durchführbarkeitsstandards erwähnt (DeGEval, 2008b; Patry & Hager, 2000). Auch in diesem Projekt mussten zugunsten der Durchführbarkeit einige Abstriche bei den Genauigkeitsstandards hingenommen werden. So wurden nur Selbstberichtdaten erhoben (Standard G4), weil der geplante Einsatz von Videovignetten unter anderem aus Zeit- und Kostengründen nicht umgesetzt werden konnte. Außerdem konnten nicht nur validierte Erhebungsinstrumente verwendet werden, sondern es kamen viele eigens konstruierte Items oder Kurzformen etablierter Skalen zum Einsatz, um den Aufwand für die Teilnehmenden möglichst gering zu halten (Standards G5 und D1). Des Weiteren konnte die ursprünglich geplante Zufallsstichprobe aus vermittelten Eltern nicht realisiert werden bzw. diese hätte nur durch erheblichen Zusatzaufwand möglicherweise erreicht werden können (Standard D1). Die resultierende selektive

Elternstichprobe schränkte die Validität ein (Standard G5). Ein weiterer Widerspruch ergab sich zwischen den Standardgruppen *Nützlichkeit* und *Fairness*, da die Evaluatoren zugunsten eines datenschutzkonformen, fairen Vorgehens bei der Stichprobenrekrutierung der Fachkräfte einen deutlich erhöhten Zeitaufwand in Kauf nahmen, was sich auf die Rechtzeitigkeit (Standard N7) und damit auf die Nützlichkeit der Evaluation auswirkte.

Die dargestellten Widersprüche verdeutlichen, dass Evaluatoren eine Position zwischen Forschung und Praxis innehaben, da sie neben wissenschaftlichen Standards auch die Anforderungen der Praxis erfüllen müssen, um nützliche Ergebnisse zu erbringen (vgl. Abschnitt 2.1.7.1). Zu den beschriebenen Kontroversen können keine Empfehlungen ausgesprochen werden, da in jeder Evaluation eine eigene Gewichtung widersprüchlicher Standards vorgenommen werden muss, die von den jeweiligen Rahmenbedingungen und Erfordernissen abhängt (vgl. DeGEval, 2008b; Stockmann, 2007). Die Frage der Gewichtung von Standards wird in Abschnitt 5.2.1 näher erörtert.

5.2 Bewertung der DeGEval-Standards

Im Folgenden soll auf Basis der im Rahmen der Meta-Evaluation gemachten Erfahrungen eine Rückmeldung zu den Evaluationsstandards der DeGEval gegeben werden. Zunächst wird die Eignung dieser Richtlinien für die Durchführung einer Meta-Evaluation beurteilt. Auch ihr Nutzen für die Planung und Durchführung von Evaluationen wird besprochen und es wird auf die in Kapitel 3 angeführten Fragestellungen Bezug genommen. Danach wird auf die Diskussion zu allgemeinen vs. feldspezifischen Standards eingegangen und es werden Vorschläge zur Überarbeitung der DeGEval-Standards unterbreitet.

5.2.1 Die DeGEval-Standards als Bewertungsinstrument bei einer Meta-Evaluation

Die Evaluationsstandards der DeGEval stellten ein sehr hilfreiches Instrument für die Durchführung der vorliegenden Meta-Evaluation dar, mussten allerdings durch das Handbuch zu den amerikanischen

Richtlinien ergänzt werden. Sie können dahingehend als umfassend bewertet werden, dass die 25 Einzelstandards alle wichtigen Bereiche abdeckten. Ihre Anwendung führte zu einer kritischen Reflexion in allen relevanten Bereichen, kein Aspekt der Evaluation blieb unreflektiert. Die DeGEval-Richtlinien erwiesen sich allerdings als zu unkonkret, um Bewertungskriterien ableiten zu können. Hierzu musste zusätzlich das JCSEE-Handbuch herangezogen werden, das deutlich umfangreichere Erläuterungen enthält und damit der Meta-Evaluatorin eine bessere Orientierungshilfe bot. Dieses Werk wird von der DeGEval auch explizit als Ergänzung zu ihren Standards benannt. Auf die geringe Ausführlichkeit der deutschen Richtlinien und auf die Notwendigkeit, ergänzend das amerikanische Handbuch zu verwenden, wird in Abschnitt 5.2.4 näher eingegangen.

Zu beachten ist, dass zur Durchführung der Meta-Evaluation eine intensive Einarbeitung in die Evaluationsstandards notwendig war, um ein umfängliches Verständnis dieser Richtlinien zu erhalten (was sich auch in der Untersuchung von Brandt, 2009, gezeigt hatte). Dieser Arbeitsschritt erforderte einen gewissen Zeitaufwand, sodass die Praktikabilität der Evaluationsstandards als Instrument für Meta-Evaluationen in Frage gestellt werden könnte. Denn womöglich werden Meta-Evaluationen nicht durchgeführt, wenn die dazu benötigten Instrumente viel Einarbeitungszeit erfordern. Dem ist allerdings entgegenzuhalten, dass einem kürzeren Instrument wie z. B. einer Checkliste wichtige Informationen fehlen würden und der Anwender kein so tiefes Verständnis von den Standards entwickeln könnte wie bei der Verwendung der DeGEval-Standards und der Handbuch-Erläuterungen. Auf Basis der vorliegenden Meta-Evaluation kann daher eine Checkliste oder andere Kurzform der Standards zur Durchführung von Evaluationsbewertungen nicht empfohlen werden, wenn diese als *alleiniges* Instrument zur Anwendung käme. Eine Checkliste als *Ergänzung* der bisherigen Standard-Werke könnte sich hingegen als nützlich erweisen. Diese Überlegung wird in Abschnitt 5.3.1 aufgegriffen.

Im Rahmen der Meta-Evaluation wurde deutlich, dass einige Informationen zum Evaluationsprojekt oder aus den Stellungnahmen mehreren Standards zuzuordnen waren. Dies zeigt, dass die Einzelstandards keine eindeutige Einordnung aller Informationen ermöglichen, was Widmer (1996) bereits für die amerikanischen Richtlinien bemerkt hatte. Auch die von Widmer postulierten Zusammenhänge zwischen einigen Standards konnten hier festgestellt werden und

äußerten sich in vielen Querverweisen. Dass die Standards die Kriterien der Unabhängigkeit und der Eindeutigkeit nicht erfüllten (worauf sie auch keinen Anspruch erheben), erwies sich jedoch nicht als problematisch für die vorliegende Evaluationsbewertung, sondern machte vielmehr deutlich, dass einige Aspekte einer Evaluation aus verschiedenen Blickwinkeln als relevant zu betrachten sind.

Als etwas problematisch lässt sich hingegen die Tatsache einordnen, dass die Orientierung an den DeGEval-Standards und dem JCSEE-Handbuch zu einer eher partizipativ ausgerichteten Bewertung geführt hatte, die nicht ganz zu dem bei der Evaluation verfolgten auftraggeberorientierten Vorgehen passte (vgl. Abschnitt 4.2.1). Insbesondere bei den Nützlichkeitsstandards wurden Hinweise dazu vermisst, wieviel Einbezug weiterer Beteiligter bei einem auftraggeberorientierten Ansatz stattfinden sollte und als angemessen zu bewerten ist (Abschnitt 5.1.3). In den DeGEval-Standards findet sich eine Bemerkung, wonach die Standards unabhängig vom angewendeten Evaluationsansatz Gültigkeit beanspruchen. Dieser Hinweis deckt sich allerdings nicht mit der Wahrnehmung der Standards als eher partizipativ ausgelegte Richtlinien in der vorliegenden Meta-Evaluation. Natürlich ist zu beachten, dass diese Meta-Evaluation Fallstudiencharakter hat (vgl. Abschnitt 5.3.2) und es durchaus möglich ist, dass andere Nutzer der Standards diesen keine partizipative Orientierung zuschreiben. Der dürftige Literaturbestand zu den DeGEval-Standards erbringt hierzu leider kaum ergänzende Informationen. Die in Abschnitt 2.2.5.4 geschilderte Äußerung eines Befragten, wonach der Umgang mit dem Standard N4 schwierig sei, da man den Auftraggeber klar berücksichtige, bei anderen Adressaten aber Schwierigkeiten sehe (DeGEval, 2004a), lässt keinen eindeutigen Schluss darauf zu, ob sich lediglich die Umsetzung dieses Standards herausfordernd gestaltet, oder ob seine Formulierung zu wenig Hilfestellung für die Umsetzung bietet. Daher soll an dieser Stelle keine Empfehlung ausgesprochen, sondern vielmehr eine Diskussion dazu angestoßen werden, ob die Standards der DeGEval tatsächlich für verschiedene Evaluationsansätze gleichermaßen Gültigkeit besitzen. Diesbezüglich sei angemerkt, dass womöglich gerade auftraggeberorientierte Evaluationen einen Ansatz darstellen, bei dem Evaluatoren viel Hilfestellung benötigen würden, um alle Stakeholder angemessen einzubeziehen.

Schließlich sei auf die Gewichtung der Evaluationsstandards eingegangen. Es werden keine Angaben gemacht, welche Standards als

wichtiger zu bewerten und damit höher zu gewichten wären als andere, da dies vom jeweiligen Einzelfall abhängt (vgl. Stockmann, 2007). Zwar würde eine vorab festgelegte Gewichtung der Standards dem Anwender noch mehr Orientierung bieten, da die Standards dann in eine Rangfolge gebracht und prioritär behandelt werden könnten. Jedoch könnten auch aus der vorliegenden Meta-Evaluation keine Empfehlungen zur Gewichtung gegeben werden, die sich auf andere Studien übertragen ließen, da die Gewichtung von zu vielen Faktoren wie den Rahmenbedingungen oder auch dem Evaluationsansatz abhängt. Daher wird im Einklang mit der DeGEval (2008b) dazu geraten, die Standards bei Evaluationsbewertungen nicht rein schematisch anzuwenden und nur zu untersuchen, wie viele Standards (nicht) erfüllt wurden, im Sinne eines vergleichbaren Summenwertes. Stattdessen sind die Evaluationsaktivitäten beeinflussende Faktoren wie Rahmenbedingungen oder der Evaluationsansatz zu berücksichtigen und es sollte beurteilt werden, ob in Anbetracht dieser Faktoren die einzelnen Standards angemessen eingehalten wurden.

Empfehlung 24: Eine quantitative Bewertung von Evaluationen im Sinne von Summenwerten, bei denen lediglich bewertet wird, wie viele Standards (nicht) erfüllt wurden, wird ebenso wie eine festgelegte Gewichtung der Standards nicht für sinnvoll erachtet. Stattdessen sollen Meta-Evaluatoren Faktoren, welche die Evaluationsaktivitäten beeinflussen (z. B. Rahmenbedingungen oder der gewählte Evaluationsansatz) berücksichtigen und sie sollen beurteilen, ob in Anbetracht dieser Faktoren die einzelnen Standards angemessen eingehalten wurden.

Insgesamt kann festgehalten werden, dass die DeGEval-Standards, ergänzt durch die Erläuterungen des JCSEE-Handbuchs, sich in der vorliegenden Arbeit als sehr nützliches Instrument für die Durchführung einer Evaluationsbewertung erwiesen und es ermöglichten, umfassend die verschiedenen Stärken, Schwächen und Herausforderungen einer Evaluation herauszuarbeiten. Diese Erkenntnis spiegelt die Erfahrung von Lynch et al. (2003) wider und steht im Einklang mit dem Befund von Ruhe und Boudreau, der sich auf die aktuelle Version der amerikanischen Evaluationsstandards bezieht: „We found that meta-evaluation guided by the standards can illuminate various values, issues, dilemmas, tensions and hazards commonly encountered in

evaluation studies“ (Ruhe & Boudreau, 2013, S. 927). Es wird daher empfohlen, die Standards auch bei künftigen Evaluationsbewertungen zu verwenden.

Empfehlung 25: Die DeGEval-Standards, kombiniert mit dem Handbuch des JCSEE, stellen ein hilfreiches Instrument für Meta-Evaluationen dar und werden daher zur Anwendung in diesem Bereich empfohlen.

5.2.2 *Zur Verwendung der DeGEval-Standards bei der Durchführung von Evaluationen*

In der hier bewerteten Evaluation waren die Evaluationsstandards nicht als Orientierungshilfe im Evaluationsprozess verwendet worden. Die Betrachtung der dennoch eingehaltenen Standards lässt daher Rückschlüsse auf die Intuitivität der Richtlinien zu. Ein Blick auf die Übersicht in Abbildung 7 (Seite 199) zeigt, dass die Einhaltung der Genauigkeitsstandards am besten gelungen war. Auch die Fairnessstandards waren gut eingehalten worden, während bei den Nützlichkeits- und Durchführbarkeitsrichtlinien mehr Abstriche zu verzeichnen waren. Dabei ist zu beachten, dass einige Standards miteinander im Widerspruch stehen und daher gar nicht alle Richtlinien in gleichem Maße eingehalten werden können. Dass die Genauigkeitsrichtlinien am besten erfüllt wurden, bestätigt die Hypothese aus Kapitel 3: Der einschlägigen Forschung (Finn et al., 1997; Newman & Brown, 1992; Schiffler & Hübner, 2000; Scott-Little et al., 2002) zufolge bilden die Genauigkeitsstandards Normen wissenschaftlichen Arbeitens ab, die Wissenschaftlern bekannt sind und daher von diesen mit am besten eingehalten werden sollten. Auch im vorliegenden Projekt bestand das Evaluationsteam aus Wissenschaftlern, denen diese Normen geläufig waren und die sie daher weitgehend einhielten. Auch dass für die Gruppe der Fairnessstandards eine gute Einhaltung der Richtlinien gefunden wurde, stimmt mit der entsprechenden Hypothese des Kapitels 3 überein und passt zu den Ergebnissen anderer Studien (Finn et al., 1997; Newman & Brown, 1992; Widmer, 1996). Im vorliegenden Projekt wurde den hierunter fallenden Aspekten wie Datenschutz sowie Würde und Rechte der Beteiligten oberste Priorität eingeräumt, was mit einer guten Befolgung der entsprechenden Standards einherging. Die

etwas größere Anzahl an Verstößen gegen die Nützlichkeits- und Durchführbarkeitsstandards wurde ebenfalls zumindest teilweise bei anderen Studien beobachtet (Finn et al., 1997; Newman & Brown, 1992; Ruhe & Boudreau, 2013; Scott-Little et al., 2002) und kann als Hinweis darauf gesehen werden, dass nicht alle Standards gleichermaßen intuitiv eingehalten werden. Als Konsequenz kann abgeleitet werden, dass gerade diese Richtlinien noch bekannter gemacht werden sollten. Dies deckt sich mit einer Forderung im Rahmen der Revisionsbefragung der DeGEval, wonach die nicht-methodologischen Standards bekannter gemacht werden sollten (DeGEval, 2004a; vgl. Abschnitt 2.2.5.4).

Empfehlung 26: Während einigen Standards eine gewisse Intuitivität zugeschrieben werden kann, sollten insbesondere die Richtlinien zur Nützlichkeit und Durchführbarkeit von Evaluationen noch bekannter gemacht werden.

Im Theorieteil wurde berichtet, dass mitunter die Realitätsferne der Evaluationsstandards bemängelt wird (Abschnitte 2.2.5.3 und 2.2.5.4). Dieser Kritikpunkt findet anhand der vorliegenden Meta-Evaluation keine Bestätigung. Natürlich sind mit den ausdrücklich als Maximalstandards formulierten Richtlinien hohe Anforderungen verbunden, die als unrealistisch betrachtet werden können, wenn davon ausgegangen wird, dass alle Standards gleichermaßen erfüllt werden können und müssen. Darauf, dass Letzteres nicht machbar und auch nicht Ziel der Standards ist, wird in den Richtlinien selbst hingewiesen. Es gilt also, die Standards angemessen zu verstehen und in ihnen eine Orientierungshilfe dafür zu sehen, welche Aspekte bei der Evaluationsplanung und -durchführung zu beachten sind. Zudem sei angemerkt, dass auch die teilweise (vgl. DeGEval, 2004a) geforderten Minimalstandards vermutlich keine optimale Lösung darstellen würden. Denn es dürfte schwierig bis unmöglich sein, Mindeststandards, die von Evaluationen eingehalten werden sollen, so zu formulieren, dass diese für alle Evaluationsfelder und -projekte Geltung haben. Die bisherige Formulierung von Maximalstandards in Verbindung mit dem Hinweis, dass nicht alle Standards eingehalten werden können und die Nicht-Einhaltung von Standards abgewogen und begründet werden muss, wird auf Grundlage der Erfahrungen dieser Arbeit für sinnvoll erachtet.

Wenngleich in der vorgestellten Evaluationsstudie die Standards nicht verwendet wurden und erst im Rahmen der Meta-Evaluation

zum Einsatz kamen, können dennoch einige Hinweise und Empfehlungen zum Gebrauch der Richtlinien bei der Planung und Durchführung von Evaluationen gegeben werden (mit Verweis darauf, dass diese Empfehlungen nicht auf konkreter Erfahrung beruhen, sondern aus den Lernprozessen im Rahmen der Meta-Evaluation abgeleitet sind). Zunächst wird im Einklang mit anderen Autoren (Ruhe & Boudreau, 2013; Stufflebeam, 1999b) empfohlen, die Evaluationsstandards (der DeGEval, ergänzt durch das JCSEE-Handbuch) von Beginn an bei Evaluationsprojekten einzusetzen, da sie eine wertvolle Orientierungshilfe für alle Phasen einer Evaluation darstellen. In diesem Zusammenhang wird auch das funktionale Inhaltsverzeichnis der Standards, in welchem die einzelnen Standards den verschiedenen Phasen und Aufgaben einer Evaluation zugeordnet sind, positiv bewertet.

Empfehlung 27: Die Standards für Evaluation sollten von Beginn an bei Evaluationsprojekten eingesetzt werden, um deren Planung und Durchführung zu unterstützen. Die eher knapp gehaltenen DeGEval-Standards werden durch das ausführlichere Handbuch des JCSEE sinnvoll ergänzt.

Neben der Funktion einer Orientierungshilfe können die Standards auch die Kommunikation mit dem Auftraggeber erleichtern (vgl. Abschnitt 2.2.1). Denn sie liefern ein systematisches Instrument zur Besprechung kritischer Punkte und geben hilfreiche Hinweise zur Gestaltung vertraglicher Vereinbarungen. Es wird empfohlen, dass Evaluatoren die Standards zu Beginn mit dem Auftraggeber und ggf. mit weiteren Stakeholdern besprechen, eine Empfehlung, die von Ruhe und Boudreau (2013) in ähnlicher Form ausgesprochen wurde. Dabei können systematisch die Möglichkeiten und Grenzen von Evaluationen aufgezeigt werden, sodass alle Beteiligten realistische Erwartungen haben. Auch die nachfolgende Kommunikation dürfte durch die Schaffung eines einheitlichen Bezugsrahmens vereinfacht werden.

Die Erläuterungen des vorigen Abschnitts zur aufwendigen Einarbeitung, zur Gültigkeit für verschiedene Evaluationsansätze und zur Gewichtung der Standards gelten für den Gebrauch der Standards bei Evaluationsbewertungen ebenso wie für ihren Einsatz bei Evaluationsstudien und werden daher an dieser Stelle nicht erneut ausgeführt. Hinsichtlich der Gewichtung sei allerdings darauf hingewiesen, dass

eine solche, sofern dies sinnvoll erscheint, mit dem Auftraggeber und ggf. weiteren Stakeholdern gemeinsam vorgenommen werden kann.

> **Empfehlung 28**: Evaluierende sollten die Standards für Evaluation zu Beginn mit ihrem Auftraggeber und ggf. weiteren Beteiligten besprechen und dabei die Möglichkeiten und Grenzen von Evaluationen thematisieren. Mit diesen Stakeholdern kann auch eine Abstimmung in Bezug auf die Gewichtung der Standards erfolgen.

Im Rahmen der Meta-Evaluation wurde deutlich, dass nicht nur die Evaluatoren für die Einhaltung der Evaluationsstandards verantwortlich sind, sondern auch der Auftraggeber und weitere Beteiligte hierbei eine Rolle spielen können. Daher wird empfohlen, den Stakeholdern genau dies zu vermitteln.

> **Empfehlung 29**: Evaluatoren sollten den verschiedenen Stakeholdern vermitteln, dass auch diese zum Gelingen einer guten, den Standards entsprechenden Evaluation beitragen können.

5.2.3 Zur Diskussion über allgemeine vs. feldspezifische Evaluationsstandards

Im Rahmen der Revisionsbefragung der DeGEval (DeGEval, 2004a) wurde ein Diskussionspunkt deutlich, zu dem hier Stellung genommen werden soll. In dieser Studie äußerten sich einige der Befragten positiv über die Allgemeinheit der Standardformulierungen. Weit mehr Befragte wünschten sich jedoch feldspezifische Standards, also spezielle Richtlinien für verschiedene Evaluationsgebiete. Die derzeitige allgemeine Formulierung der Standards erlaubt ihre Verwendung in vielen, vielleicht sogar allen Bereichen der Evaluationsforschung, erfordert jedoch auch eine Anpassung an die jeweiligen feldspezifischen Gegebenheiten. Aufgrund der Erfahrungen im Rahmen der vorliegenden Meta-Evaluation wird die momentane Form der Standards als allgemeine Richtlinien präferiert. Dies ist zum einen der Fall, weil die Standards in dieser Form für die Evaluationsbewertung gut einsetzbar waren und nur wenige Anforderungen als nicht passend wahrgenommen wurden. Zum anderen wird, wie bereits im letzten Abschnitt anklang, die

Formulierung von Evaluationsstandards, die für alle Evaluationsfelder *und* für alle in diesen Feldern vorkommenden Evaluationsprojekte gleichermaßen Geltung besitzen, für schwierig bis unmöglich gehalten. Daher ergäbe sich vermutlich auch bei feldspezifischen Standards die Problematik, dass sie nicht für jede Evaluation in diesem Bereich passend wären, sondern in Abhängigkeit der jeweiligen Rahmenbedingungen adaptiert werden müssten. Daher erscheint der Nutzen eigener Standards für jedes Evaluationsgebiet fraglich, insbesondere, wenn er gegen den Aufwand ihrer Erarbeitung abgewogen wird. Ein Kompromiss könnte darin bestehen, dass die DeGEval-Standards als Grundlage genommen und für jeden Bereich eigens adaptiert bzw. ergänzt werden, ein Vorschlag, den auch Beywl und Speer (2003, 2004) für den Bereich der beruflichen Bildung unterbreitet haben.

Empfehlung 30: In der Diskussion um allgemeine vs. feldspezifische Evaluationsstandards wird die derzeitige allgemeine Standardformulierung präferiert. Als Kompromiss könnten die verschiedenen Evaluationsbereiche die Standards für ihre Gegebenheiten adaptieren bzw. ergänzen.

5.2.4 Vorschläge für eine Überarbeitung der DeGEval-Standards

Basierend auf dieser Meta-Evaluation können einige Vorschläge zur Überarbeitung der DeGEval-Standards generiert werden, da diese als Instrument für eine Evaluationsbewertung getestet wurden. Dies erscheint insbesondere deshalb sinnvoll, weil derzeit ein Revisionsprozess der Standards durchgeführt wird. Natürlich ist nicht zu erwarten, dass alle hier gegebenen Anregungen im Revisionsprozess direkt umgesetzt werden, da sie nur eine von vielen, sicherlich teilweise gegensätzlichen Rückmeldungen darstellen werden. Dennoch soll der Revisionsprozess mit den Erfahrungen und Erkenntnissen aus der vorliegenden Meta-Evaluation unterstützt werden.

Zunächst sei auf die geringe Ausführlichkeit der deutschen Evaluationsstandards im Vergleich zu ihrem amerikanischen Pendant eingegangen. Für die Durchführung der Meta-Evaluation mussten die DeGEval-Richtlinien um die Erläuterungen des JCSEE-Handbuchs

erweitert werden. Dies war deshalb erforderlich, weil die vergleichsweise knapp gehaltenen deutschen Standards sich als zu unkonkret erwiesen, um daraus Bewertungskriterien ableiten zu können (Abschnitt 4.2.1). Auch in der Evaluationspraxis dürfte es notwendig sein, die DeGEval-Richtlinien durch das Handbuch zu den amerikanischen Standards zu ergänzen, da Erstere zu wenig konkrete Handlungsanleitungen enthalten. Nun ist zu bemerken, dass die Verwendung ergänzender Literatur zu den DeGEval-Standards von deren Autoren vorgesehen ist (vgl. Abschnitt 2.2.2) und insbesondere das JCSEE-Handbuch ausdrücklich
erwähnt wird. Daher sind die hier getroffenen Aussagen zur Notwendigkeit, die deutschen Standards zu ergänzen, keinesfalls als Kritik zu verstehen, sondern beschreiben lediglich die Erfahrungen, die im Rahmen der Meta-Evaluation gemacht wurden.

In der Konsequenz bedeutet das, dass die DeGEval-Standards nicht als alleiniges Instrument für die Durchführung von Meta-Evaluationen ausreichen, und dass Ähnliches auch für ihren Einsatz bei der Planung und Durchführung von Evaluationen gelten dürfte. In beiden Anwendungsfällen erscheint die Kombination der deutschen Richtlinien mit dem amerikanischen Handbuch erforderlich. Sofern der DeGEval diese Konstellation weiterhin ausreicht und sie mit dem Geltungsbereich ihrer Standards zufrieden ist, sind diesbezüglich keine Empfehlungen auszusprechen. Für den Fall jedoch, dass die DeGEval mit ihren Standards ein Werk anstrebt, welches für sich genommen Evaluationen und Meta-Evaluationen anleiten kann, wird empfohlen, die Standards zu einem ausführlichen Handbuch weiterzuentwickeln. Dazu müssten die Erläuterungen zu den einzelnen Standards deutlich erweitert werden, um konkretere Orientierungshilfen zu bieten. Denkbar sind z. B. Ratschläge in Form von Dos and Don'ts, also Handlungsempfehlungen sowie Hinweise auf häufige Fehler im Zusammenhang mit jedem Standard, wie sie im Handbuch des JCSEE enthalten sind. Eine solche Konkretisierung müsste sich nicht auf bestimmte Evaluationsgebiete beziehen, würde also die Allgemeinheit der Standards (im Sinne der Passung für alle Evaluationsbereiche, vgl. Abschnitt 5.2.3) nicht beeinträchtigen. Diese ausführlicheren Erläuterungen könnten durch Fallbeispiele (ebenfalls analog zum JCSEE-Handbuch) sinnvoll ergänzt werden, welche die jeweiligen Standard-Anforderungen veranschaulichen und damit zu einem besseren Verständnis von deren (Nicht-) Einhaltung beitragen würden.

Die Erarbeitung eines umfangreichen Handbuchs würde allerdings einen hohen Arbeitsaufwand bedeuten, sodass eine Entscheidung der DeGEval gegen eine solche ausführliche Version verständlich wäre. Dennoch gilt es einige weitere Aspekte zu bedenken, die für ein Handbuch zu den deutschen Evaluationsstandards sprechen. Zunächst ist die Handhabung zweier Werke (der DeGEval-Standards und des JCSEE-Handbuchs) als weniger praktikabel einzuschätzen als die Verwendung eines Werkes. Das Vorliegen eines umfassenden Handbuchs zu den deutschen Richtlinien könnte daher deren Nutzung in der Praxis erhöhen. Auch sind die amerikanischen und die deutschen Standards nicht identisch, sodass das Handbuch zum DeGEval-Standard N2 keine Orientierung bietet, weil er dort fehlt. Zudem umfassen die deutschen Richtlinien weniger Standards als die amerikanischen, weil dort mehrfach zwei der JCSEE-Standards zusammengefasst wurden. Zwar existiert eine Transformationstabelle, welche eine Zuordnung der deutschen zu den amerikanischen Einzelstandards enthält, was das Nachschlagen der passenden Standard-Erläuterungen im JCSEE-Handbuch erleichtert. Das Vorliegen eines ausführlichen Handbuchs zu den DeGEval-Standards würde jedoch eine deutlich praktikablere Lösung darstellen. Schließlich sei erwähnt, dass die Fallbeispiele des JCSEE-Handbuchs aus dem amerikanischen Raum stammen. Wenngleich sie zur Veranschaulichung der einzelnen Standards ausreichen, dürften Fallbeispiele aus Deutschland für die Anwender der deutschen Standards noch besser nachvollziehbar sein und eine höhere Praxisrelevanz aufweisen.

Aus den angeführten Gründen wird dazu geraten, die DeGEval-Standards in Richtung eines ausführlicheren Handbuchs zu erweitern. Diese Empfehlung spiegelt den Befund der Revisionsbefragung der DeGEval wider, wonach die Entwicklung eines umfangreicheren Handbuchs die am häufigsten gewünschte Revisionsform darstellte (DeGEval, 2004b).

Empfehlung 31: Die DeGEval-Standards sollten zu einem umfassenden Handbuch mit ausführlicheren Erläuterungen und konkreten Handlungsempfehlungen sowie Fallbeispielen weiterentwickelt werden, um ihre Relevanz und Nutzung zu erhöhen.

Falls ein Handbuch zu den DeGEval-Standards realisiert werden sollte, könnte überlegt werden, neben dieser ausführlichen Version die bisherige kurze Form beizubehalten. Letztere könnte z. B. bei der Kommunikation der Standards an den Auftraggeber und andere Stakeholder Verwendung finden, da für diesen Zweck ein umfassendes Handbuch womöglich nicht geeignet wäre. Für die lange Version wäre ein etwas anderer Aufbau als im amerikanischen Werk zu empfehlen, da in jenem etliche Punkte doppelt genannt werden. Das JCSEE-Handbuch enthält nach dem Namen und der Kurzbeschreibung des Standards eine Übersicht (Erklärungen zum Standard), gefolgt von Richtlinien (Handlungsempfehlungen), häufigen Fehlern und den Fallbeispielen. Da in der Übersicht bereits einige Handlungsanleitungen gegeben werden, kommt es zu vielen Überschneidungen mit dem Punkt „Richtlinien". Für die mögliche Handbuchversion der DeGEval-Standards wird daher empfohlen, nach dem Standardnamen und der Kurzbeschreibung eine Übersicht oder Erklärung folgen zu lassen, die lediglich den Standard beschreibt und noch keine Handlungsempfehlungen enthält. In den sich anschließenden Dos and Don'ts können Tipps gegeben und häufige Fehler benannt werden, bevor Fallbeispiele den Standard abrunden[10].

Schließlich seien zwei weitere Vorschläge unterbreitet, wie die DeGEval-Standards sinnvoll ergänzt werden könnten. Zum einen sollte in den Standards deutlicher darauf hingewiesen werden, dass nicht allein die Evaluierenden die Einhaltung der Standards zu verantworten haben, sondern auch der Auftraggeber und ggf. andere Beteiligte dazu ihren Beitrag leisten können (vgl. Abschnitt 5.1.3.2). In der bisherigen Version der deutschen Standards fand dieser Aspekt lediglich im Rahmen des Standards F5 (Offenlegung der Ergebnisse) Erwähnung. Da er sich jedoch auf alle Standards beziehen sollte, wäre ein entsprechender

[10] Im Rahmen des laufenden Revisionsprozesses hat die DeGEval eine Entwurfsfassung für die neue Version der Standards erarbeitet. In diesem Papier wurde der hier vorgeschlagene Hinweis dahingehend umgesetzt, dass dort die „Begründung" des Standards von den „Umsetzungshinweisen" getrennt aufgeführt wird. Die Entwurfsfassung wurde der Autorin im September 2015 im Rahmen der 18. DeGEval-Jahrestagung zugänglich gemacht. Da sie die genannte Empfehlung zu diesem Zeitpunkt, unabhängig von der neuen Entwurfsfassung, bereits generiert hatte, wird sie hier dennoch ausgesprochen, auch wenn sie offenbar schon umgesetzt wurde.

Hinweis bereits im Einleitungstext zu den Standards angemessen[11]. Zum anderen wird in der momentanen Version der Standards lediglich im Vorwort auf die „Empfehlungen für Auftraggebende von Evaluationen“ (DeGEval, 2007) hingewiesen. Diese Broschüre enthält neben Ratschlägen für Auftraggebende zur Vorbereitung und Ausschreibung einer Evaluation auch Hinweise zur Begleitung der Evaluation und zur Verbreitung und Umsetzung der Ergebnisse. Somit erscheint sie als hilfreiches Instrument z. B. bei der Verständigung auf bestimmte Vorgehensweisen und auf die Evaluationsstandards und könnte die Kommunikation zwischen Evaluatoren und Auftraggebern unterstützen. Wenn Evaluierende ihrem Auftraggeber diese Empfehlungen aushändigen, verfügt dieser zudem über eine fundierte Richtlinie für weitere Evaluationsprojekte. Daher sollte in den DeGEval-Standards deutlicher auf die „Empfehlungen für Auftraggebende“ hingewiesen und Evaluatoren empfohlen werden, diese Broschüre mit ihrem Auftraggeber zu besprechen.

5.3 Kritische Reflexion der angewandten Methode

In diesem Abschnitt wird die verwendete Methode kritisch reflektiert. Dabei werden zunächst die Stärken und Schwächen der vorliegenden Meta-Evaluation analysiert, dann folgt eine Erörterung des Geltungsbereichs dieser Arbeit und der ausgesprochenen Empfehlungen. Nachfolgend werden weitere methodische Überlegungen angeführt, bevor ein abschließendes Fazit zur vorliegenden Studie einschließlich Hinweisen zu Forschungsimplikationen die Diskussion abrundet.

[11] Im Rahmen des laufenden Revisionsprozesses hat die DeGEval eine Entwurfsfassung für die neue Version der Standards erarbeitet. In diesem Papier ist der hier vorgeschlagene Hinweis im einleitenden Teil enthalten. Die Entwurfsfassung wurde der Autorin im September 2015 im Rahmen der 18. DeGEval-Jahrestagung zugänglich gemacht. Da sie die genannte Empfehlung zu diesem Zeitpunkt, unabhängig von der neuen Entwurfsfassung, bereits generiert hatte, wird sie hier dennoch ausgesprochen, auch wenn sie offenbar schon umgesetzt wurde.

5.3.1 Stärken und Schwächen der vorliegenden Meta-Evaluation

Wie alle wissenschaftlichen Studien weist auch die hier vorgestellte Meta-Evaluation methodische Stärken und Schwächen auf. Diese gründlich zu analysieren ist als besonders wichtig zu erachten, weil der bisherige Forschungsstand zur Methodik bei Evaluationsbewertungen als dürftig zu bewerten ist und mit der vorliegenden Arbeit das Anliegen verbunden war, einen Beitrag zur diesbezüglichen fachlichen Diskussion zu leisten.

In der vorliegenden Studie wurde eine Methode gewählt, die sich an der von Widmer (1996) orientierte und sich während des Bewertungsprozesses weiterentwickelte. So waren die Stellungnahmen mit Beteiligten nicht von vornherein geplant gewesen, sondern erwiesen sich erst im Verlauf der Bewertung als notwendig bzw. nützlich. Interessanterweise ergaben sich dadurch im Nachhinein mehr Überschneidungspunkte mit Widmers Methode, als zu Beginn der Studie geplant gewesen waren. Im Folgenden werden die einzelnen Schritte der angewandten Methodik kritisch beleuchtet.

Der erste Schritt bestand in der intensiven Einarbeitung in die DeGEval-Standards und in das JCSEE-Handbuch. Diese war zeitaufwendig, aber erforderlich, um mit den Standards vertraut zu werden und Bewertungskriterien ableiten zu können. Aus diesem Grund wird auch der Standpunkt vertreten, dass Checklisten bei Meta-Evaluationen zwar ergänzend zum Einsatz kommen können, um den Bewertungsprozess zu strukturieren und zu verhindern, dass wichtige Punkte übersehen werden (vgl. Yarbrough et al., 2011). Dies setzt jedoch voraus, dass der Nutzer sich bereits mit den Standards befasst und ein Verständnis für ihre Anforderungen entwickelt hat. Ein alleiniges Instrument für Evaluationsbewertungen können Checklisten daher nicht darstellen, weil sie dem Nutzer zu wenige Informationen bieten, um die Standardanforderungen verstehen und überprüfen zu können.

Die Tatsache, dass in der vorliegenden Studie neben den deutschen Evaluationsstandards auch die amerikanischen Richtlinien verwendet wurden, könnte kritisch gesehen werden, weil dadurch der Geltungsbereich der ausgesprochenen Empfehlungen verschwimmen könnte – beziehen diese sich auf die deutschen oder auf die amerikanischen Standards oder auf beide? Dieser Einwand kann allerdings entkräftet werden, da sich die beiden Richtlinien einerseits sehr ähnlich sind und andererseits die deutschen Standards absichtlich nicht neu formuliert

oder umstrukturiert wurden, um eine Vergleichbarkeit mit anderen Werken (darunter dem Handbuch zu den amerikanischen Standards) zu ermöglichen (vgl. Abschnitt 2.2.2).

Im nächsten Schritt wurden für jeden Standard Bewertungskriterien erarbeitet sowie Kategorien zur Beurteilung der Einhaltung der Standards erstellt. Die Bewertungskriterien wurden aus den in den deutschen und amerikanischen Richtlinien genannten Anforderungen abgeleitet. Einige Anforderungen aus dem JCSEE-Handbuch wurden nicht in den Kriterienkatalog mit aufgenommen, da sie für das vorliegende Evaluationsprojekt nicht zutrafen. Diese wurden auch im Ergebnisteil dieser Arbeit nicht angeführt (Abschnitt 4.2.2). Dieses leicht selektive Vorgehen kann subjektiv anmuten und schränkt die Nachvollziehbarkeit der angewandten Methode etwas ein. Es wurde verfolgt, weil die Darstellung sämtlicher in den Standards genannten Anforderungen den Rahmen dieser Arbeit überstiegen hätte. Es musste daher ein Kompromiss zwischen vollständiger Darlegung und Nachvollziehbarkeit einerseits und Darstellungsökonomie andererseits eingegangen werden. An dem letztendlich gewählten Vorgehen könnte kritisiert werden, dass bei der Festlegung der nicht zutreffenden Kriterien womöglich auch solche weggelassen wurden, die von der Evaluation nicht eingehalten wurden, um eine bessere Bewertung der Evaluation zu erhalten. Dass dies nicht der Fall war, wird dadurch deutlich, dass bei den verschiedenen Standards etliche Abzüge berichtet wurden, was zeigt, dass eine ausgewogene und faire Bewertung und Darstellung stattfand. Dennoch erscheint das Vorgehen im Hinblick auf die Nachvollziehbarkeit verbesserungswürdig. Es wurde dadurch erschwert, dass kein Kriterienkatalog vorlag, sondern eigens zusammengestellt werden musste. Für die Zukunft wäre es hilfreich, einen vollständigen Kriterienkatalog zu erarbeiten, der für jeden Standard *alle* zu erfüllenden Anforderungen enthält (ähnlich der Methodik von Stufflebeam, 2001b, vgl. Abschnitt 2.2.5.2). Diese können dahingehend bewertet werden (einschließlich einer Begründung der Bewertung), ob sie eingehalten wurden oder nicht. Je nachdem, wie viele Anforderungen eingehalten wurden, kann dann der Standard als mehr oder weniger erfüllt beurteilt werden. Wichtig ist jedoch, dass neben der Bewertung einer Anforderung als *(nicht) eingehalten* auch das Urteil *nicht zutreffend* möglich ist, um eine Relativierung des Urteils und eine gerechte Bewertung zu erhalten. In diesem Zusammenhang sei auch

erwähnt, dass die quantitative Bewertung von Evaluationen mittels Summenscores, wie Stufflebeam (1999a, 2000) sie in seiner Checkliste vorschlägt, nicht für sinnvoll erachtet wird (vgl. Abschnitt 5.2.1). Denn einerseits müssen die Urteile stets an den zur Verfügung stehenden Ressourcen und dem Kontext, in dem die Evaluation stattfindet, relativiert werden. Andererseits ist die von Studie zu Studie verschiedene Gewichtung der Standards zu berücksichtigen, sodass ohnehin kein simpler Vergleich zwischen mehreren Summenscores vorgenommen werden kann.

Die vier Bewertungskategorien (*Standard erfüllt/überwiegend erfüllt/teilweise erfüllt/nicht erfüllt*) erwiesen sich als angemessen für die vorliegende Meta-Evaluation, wobei die Kategorie *nicht erfüllt* nicht vergeben wurde. Für andere Evaluationsbewertungen könnten jedoch andere Kategorien passender sein. Wenn es z. B. erforderlich ist, im unteren Bereich die Bewertungen mehr zu differenzieren, könnte zwischen *teilweise erfüllt* und *nicht erfüllt* noch die Stufe *überwiegend nicht erfüllt* eingesetzt werden (ähnlich wie bei Widmer, 1996). Wie im Methodenteil bereits angemerkt wurde, kann der Vergabe der Urteile *überwiegend erfüllt* und *teilweise erfüllt* eine gewisse Subjektivität unterstellt werden, da die jeweils erforderlichen Verstöße nicht näher operationalisiert wurden (Abschnitt 4.2.1). Dies schränkt zwar die Nachvollziehbarkeit der Methodik ein, jedoch erscheint eine numerische Festlegung der Schwere eines Verstoßes schlicht nicht sinnvoll (und wurde auch z. B. bei Widmer, 1996, nicht vorgenommen). Stattdessen wurde hier versucht, die Bewertung ausführlich zu begründen und damit ein dennoch möglichst nachvollziehbares Urteil zu liefern. Zur Reduzierung der Subjektivität wurde zudem die konservativere Bewertung gewählt, wenn das Urteil zwischen zwei Kategorien schwankte.

Es folgten der erste Bewertungsdurchgang und daran anschließend die Erhebung der Stellungnahmen bei den Stakeholdern. Diese hatten sich zuvor als erforderlich herausgestellt, um Informationslücken zu schließen sowie verschiedene Perspektiven bei der Bewertung berücksichtigen zu können und dadurch letztendlich ein weniger subjektives Urteil zu erlangen. In der Tat erwiesen sich die Stellungnahmen als sehr hilfreich, da sie wichtige, ergänzende Daten zu fast allen Standards erbrachten. Lediglich zu den Genauigkeitsstandards lieferten sie kaum einen Beitrag, was damit zu begründen ist, dass dort Aspekte wie

eine genaue Beschreibung im Bericht oder die Qualität der Datenanalyse zu bewerten sind. Die Meta-Evaluatorin hatte in den Stellungnahmen nicht gezielt nach solchen Gesichtspunkten gefragt, da sie diese einerseits anhand der ihr vorliegenden Informationen selbst hinreichend beurteilen konnte und andererseits von den meisten Stakeholdern keine differenzierte Bewertung des wissenschaftlichen Arbeitens zu erwarten war. Hinzu kam, dass der Abschlussbericht zum Zeitpunkt der Erhebung der Stellungnahmen noch nicht fertiggestellt und verbreitet gewesen war, sodass dazu seitens der Stellungnehmer keine Aussagen hätten getroffen werden können. Die Stellungnahmen auf die Zeit nach der Veröffentlichung des Abschlussberichts zu verschieben, war aus zeitlichen Gründen nicht möglich. Zudem bedeuteten die Stellungnahmen ohnehin Zeitaufwand für die Befragten, der noch deutlich vergrößert worden wäre, wenn sie zusätzlich den umfangreichen Abschlussbericht hätten lesen müssen, um diesen bewerten zu können.

Hinsichtlich der Stellungnahmen ist zu bemerken, dass die Vertreter der Gruppen der Eltern und Elternbegleiter womöglich eine positive Selektion darstellten. Denn es wurden bei den Eltern Personen ausgewählt, die sowohl am Prätest als auch am Posttest teilgenommen hatten. Bei den Elternbegleitern war ein Kriterium, dass sie Eltern zur Teilnahme an der Befragung vermittelt haben sollten. Beide Bedingungen wurden gestellt, um Personen für die Stellungnahmen zu gewinnen, die möglichst viele Erfahrungen mitteilen konnten. Allerdings wurden dadurch diejenigen Personen ausgeschlossen, zu denen der Kontakt im Verlauf der Befragung abgebrochen war – also Eltern, die nicht am Posttest teilgenommen hatten und Elternbegleiter, die keine Eltern vermittelt hatten. Dafür sind vielfältige Gründe denkbar, von denen die wahrscheinlichsten sind, dass die Eltern zum Posttestzeitpunkt nicht mehr erreichbar waren und dass die Elternbegleiter keine Eltern zur Vermittlung gefunden hatten. Allerdings ist es zumindest möglich, dass bei diesen Personen eine gewisse Unzufriedenheit oder auch ein Misstrauen gegenüber der Evaluation eine Rolle dafür gespielt haben könnten, dass sie sich nicht weiter an der Studie beteiligten. In einer Stellungnahme hätten diese Menschen womöglich mehr Kritikpunkte geäußert als dies bei der tatsächlichen, potentiell etwas positiv selektierten Stichprobe der Fall war. Allerdings stand bei der Auswahl der Stellungnehmer nicht ihre Repräsentativität im Vordergrund, sondern die Fülle an Informationen, die sie vermutlich würden mitteilen

können (vgl. Abschnitt 4.2.1). Zudem wurde nach praktischen Gesichtspunkten vorgegangen: Es sollten Personen angesprochen werden, die erreichbar und außerdem zu einer Stellungnahme bereit waren, also eine gewisse Offenheit aufwiesen und den Zeitaufwand auf sich nehmen würden. Bei den eben erwähnten Personen, die sich nicht fortdauernd an der Evaluation beteiligt hatten, war hingegen von einer schlechten Erreichbarkeit und/oder geringen Bereitschaft, sich für eine Stellungnahme im Rahmen der Meta-Evaluation zur Verfügung zu stellen, auszugehen. An diesem Beispiel wird deutlich, dass auch bei der Meta-Evaluation zwischen Genauigkeit (im Sinne einer repräsentativen Stichprobe) und Durchführbarkeit (im Sinne einer erreichbaren Stichprobe) abgewogen werden musste.

Der Vertreter des Evaluationsteams bei den Stellungnahmen war gezielt so ausgewählt worden, dass seine Stellungnahme zwei möglichen Verzerrungsursachen seitens der Meta-Evaluatorin entgegenwirken konnte. So handelte es sich um eine Person aus dem DJI, im Gegensatz zur Meta-Evaluatorin, die der FAU angehörte. Dadurch sollte ein möglicher Fokus der Meta-Evaluatorin auf das von ihr vorwiegend behandelte Modul 3 der Evaluation verringert werden. Zudem verfügte die stellungnehmende Person über mehr Berufserfahrung als die Meta-Evaluatorin, sodass auch ein diesbezüglicher Bias reduziert werden konnte.

Ein weiterer möglicher Kritikpunkt an den Stellungnahmen betrifft ihre Auswertung. Diese erfolgte inhaltsanalytisch und wurde durch die Meta-Evaluatorin vorgenommen. Auf eine Überprüfung dieser Auswertung durch einen unabhängigen zweiten Rater wurde aus zeitlichen Gründen verzichtet, außerdem waren keine Ressourcen für einen solchen Schritt vorgesehen. Auch hier musste somit ein Kompromiss zwischen Genauigkeit und Durchführbarkeit geschlossen werden. Dass dieser in Richtung Durchführbarkeit ausfiel, ist damit zu begründen, dass es sich nicht um offenes, qualitatives Material handelte, bei dem eine Zweitkodierung zur Absicherung eher notwendig gewesen wäre, sondern um Antworten auf einen Leitfaden, sodass eine valide Zuordnung auch von der Meta-Evaluatorin allein vorgenommen werden konnte (vgl. Abschnitt 4.2.1).

Insgesamt kann festgehalten werden, dass die Stellungnahmen zwar einigen Aufwand für Meta-Evaluatorin und Stellungnehmer bedeuteten, sich jedoch als sehr wertvoll für die Evaluationsbewertung erwiesen.

An die Erhebung und Auswertung der Stellungnahmen schloss sich der zweite Bewertungsdurchgang an, in dem, ergänzend zu den übrigen Daten, die Informationen aus den Stellungnahmen mit einbezogen wurden. Nachfolgend wurde die detailliert schriftlich festgehaltene Bewertung noch an den Auftraggeber und ein Mitglied des Evaluationsteams gegeben sowie deren Rückmeldung erbeten und schließlich eingearbeitet. Dieses Vorgehen wurde als sehr nützlich empfunden, um fehlende Informationen zu ergänzen, die eigene Bewertung abzusichern und somit die Subjektivität des Urteils der Meta-Evaluatorin zu reduzieren. Es könnte kritisiert werden, dass lediglich dem Auftraggeber und dem Mitglied des Evaluationsteams die Möglichkeit zur Rückmeldung gegeben wurde, den weiteren Stakeholdern hingegen nicht. Eine Rückmeldung auch von anderen Stakeholdern wäre zwar sehr wünschenswert gewesen und hätte weitere hilfreiche Informationen erbringen können. Jedoch hätte ein solcher Schritt einen erheblichen zeitlichen Aufwand für die Stellungnehmer und damit eine beträchtliche Verzögerung für vorliegende Arbeit bedeutet, sodass entschieden wurde, mit dem Auftraggeber und dem Mitglied des Evaluationsteams die beiden Stakeholder auszuwählen, die über die meiste Einsicht in den Evaluationsprozess verfügten (vgl. Abschnitt 4.2.1).

Im Rahmen der zweiten Stellungnahmen wurde eine Diskrepanz zwischen dem auftraggeberorientierten Ansatz der Evaluation und der eher partizipativ ausgerichteten Bewertung anhand der Standards deutlich, die an anderer Stelle diskutiert wurde (Abschnitt 5.2.1). Im vorliegenden Fall ist diese Diskrepanz nicht als gravierend zu bewerten, da sie lediglich bei den beiden Stellungnehmern für leichte Irritationen sorgte und die Meta-Evaluation in erster Linie dem Erkenntnisgewinn dient. Es sind jedoch Situationen denkbar, in denen eine Meta-Evaluation direktere Auswirkungen auf verschiedene Beteiligte hat und daher von diesen als unpassend wahrgenommene Bewertungskriterien für Unmut sorgen könnten. Für solche Fälle wird empfohlen, die Bewertungskriterien, die im Rahmen einer Meta-Evaluation zur Anwendung kommen sollen, vorab mit den Beteiligten zu besprechen, eine Empfehlung, die auch Thomas Widmer (1996) ausgesprochen hat.

Empfehlung 32: Die Bewertungskriterien, die bei einer Meta-Evaluation zum Einsatz kommen, sollten vorab mit den Beteiligten besprochen werden.

5.3.2 Geltungsbereich

Weiterhin soll auf den Geltungsbereich der vorliegenden Arbeit und damit auch der hier ausgesprochenen Empfehlungen eingegangen werden. So ist kritisch zu bemerken, dass es sich um eine Fallstudie handelt, bei der die im Rahmen *einer* Evaluation gemachten Erfahrungen analysiert wurden. Dieser Ansatz wurde gewählt, um eine sehr detaillierte und genaue Untersuchung der betreffenden Evaluation zu erhalten, die deshalb möglich war, weil die Meta-Evaluatorin an der Evaluation selbst beteiligt gewesen war. Demgegenüber können bei synthetischen Analysen zwar mehrere Evaluationen einbezogen, jedoch häufig nicht so detailliert beurteilt werden, weil den Meta-Evaluatoren Informationen für die Bewertung fehlen (vgl. Abschnitt 5.3.3). Da bereits einige als Evaluationssynthese durchgeführte Meta-Evaluationen existieren, erschien die Ergänzung durch den in vorliegender Arbeit verfolgten Ansatz sinnvoll, auch wenn Einschränkungen bezüglich des Geltungsbereichs denkbar sind. Denn es könnte sein, dass die Empfehlungen nur für Evaluationen mit vergleichbaren Rahmenbedingungen passend sind, also für ähnlich umfangreiche Evaluationsprojekte mit vielen Beteiligten und umfassenden Datenerhebungen; für Programmevaluationen; für Evaluationen, die einem auftraggeberorientierten Ansatz folgen; für Studien, die im Bereich der Familienbildung durchgeführt werden. Um die begrenzte Generalisierbarkeit zu erhöhen, wurden die Befunde anderer Autoren herangezogen und mit den Erkenntnissen der vorliegenden Arbeit verknüpft. Einige der hier gegebenen Empfehlungen stellen auch lediglich Bekräftigungen bereits vorhandener Ratschläge (z. B. der Standards) dar, da jene im Rahmen dieser Meta-Evaluation getestet wurden. Zudem ist darauf hinzuweisen, dass in der vorliegenden Arbeit *Empfehlungen* ausgesprochen und keine unumstößlichen, umfassend belegten Erkenntnisse postuliert werden, sodass der Geltungsbereich auch in der gewählten, zurückhaltenden Formulierung Berücksichtigung findet. Ziel dieser Arbeit konnte auch nicht sein, Empfehlungen zu erarbeiten, die für jedes zukünftige Evaluationsprojekt zutreffen würden (Kapitel 3). Die berichteten Erfahrungen und Herausforderungen sowie die generierten Ratschläge können jedoch sicher zumindest für einige künftige Evaluationen nützlich sein. Zudem sind weitere Studien wünschenswert, die analog zu vorliegender Arbeit die Erfahrungen von Evaluationsprojekten systematisch analysieren und Empfehlungen aussprechen. Denn die in Evaluationen gemachten Erfahrungen sollten zur Optimierung

künftiger Projekte genutzt werden, was bislang jedoch kaum gemacht wird (vgl. Beywl, 2001; Spiel et al., 2012). Das langfristige Ziel würde in einem deutlich erweiterten Fundus an Erfahrungen und Ratschlägen bestehen, auf den Evaluatoren künftig zurückgreifen könnten und der dann auf einer breiteren Grundlage an Evaluationen basieren würde. Damit wäre es auch möglich, solche Empfehlungen, die häufig genannt werden und daher für die meisten Evaluationen gelten dürften, von jenen zu unterscheiden, die nur bei Vorliegen gewisser Rahmenbedingungen oder in bestimmten Evaluationsbereichen zum Tragen kommen. Erstere dürften sich mit den gängigen Evaluationsstandards decken und könnten diese um die bislang vermissten, direkteren Handlungsempfehlungen ergänzen. Letztere könnten die mitunter gewünschte Konkretisierung der Standards für verschiedene Evaluationsfelder unterstützen (vgl. Abschnitt 0).

Empfehlung 33: Es wird geraten, weitere Meta-Evaluationen durchzuführen, in denen die in Evaluationen gemachten Erfahrungen systematisch analysiert und Empfehlungen erarbeitet werden.

Die im Theorieteil geschilderte Feststellung, dass bei vielen Meta-Evaluationen die eingesetzte Methode nur unzureichend beschrieben wird, gibt Anlass zu einer weiteren Empfehlung. Denn bei der vorliegenden Studie standen dadurch nur wenige Anregungen für die Wahl der eigenen Methode zur Verfügung. Daher sollten künftige Meta-Evaluatoren das von ihnen angewandte Vorgehen detaillierter beschreiben, als dies bisher der Fall ist. Die eingesetzten Bewertungskriterien sollten benannt und begründet werden. Weiterhin sollte erläutert werden, ob es sich um eine interne oder externe Meta-Evaluation handelt und in welcher Form die Bewertung erfolgt (z. B. ein oder zwei Rater; zugrunde liegende Dokumente; Erhebung von Stellungnahmen). Solche grundlegenden methodischen Hinweise sollten auch in kürzeren Publikationen in Fachzeitschriften Platz finden können, nicht nur in umfangreichen Meta-Evaluationsberichten. Idealerweise sollte die angewendete Methode auch kritisch reflektiert werden, um Stärken und Schwächen benennen zu können. Eine genauere Schilderung der Methode würde in vielfältigeren und besseren Orientierungsmöglichkeiten für künftige Meta-Evaluatoren resultieren.

Empfehlung 34: Die in Meta-Evaluationen angewandte Methode sollte detaillierter beschrieben werden, um anderen Meta-Evaluatoren eine Orientierung zu bieten.

5.3.3 Weitere methodische Überlegungen

Bei der dargestellten Evaluationsbewertung handelte es sich um eine interne Meta-Evaluation, die durch ein Mitglied des Evaluationsteams vorgenommen wurde. Während Objektivität bei Bewertungsprozessen ohnehin schwer erreichbar scheint, wird sie bei internen Meta-Evaluationen zusätzlich dadurch erschwert, dass hier zumindest teilweise eine Bewertung der eigenen Arbeit erfolgt, was einen Interessenkonflikt bedeuten kann (Yarbrough et al., 2011). Im vorliegenden Projekt bestand die erste Maßnahme, die gegen ein allzu subjektives Urteil ergriffen wurde, darin, anhand von Kriterien (den Evaluationsstandards) eine *systematische* Analyse und Bewertung vorzunehmen. Doch auch bei der Verwendung von anerkannten Bewertungskriterien bleibt die Einstufung der Evaluation ein subjektiver Prozess (vgl. Widmer, 1996). Es wurde daher beschlossen, eine möglichst detaillierte Beschreibung der Bewertungskriterien sowie der Projektdaten und Informationen aus den Stellungnahmen zu liefern, um die Einschätzung, inwieweit ein Standard als erfüllt betrachtet wurde, nachvollziehbar zu machen. Zudem wurde im Zweifel die konservativere Bewertung gewählt. Weiterhin ermöglichte die Erhebung der Stellungnahmen den Einbezug verschiedener Perspektiven und damit ein Gegengewicht zur Einschätzung der Meta-Evaluatorin. Dies reduzierte die Subjektivität ihres Urteils ebenso wie die zweiten Stellungnahmen zur verschriftlichten Evaluationsbewertung. Während also die Subjektivität der Bewertung den zentralen Kritikpunkt an der vorliegenden Studie darstellen dürfte, wurden doch etliche Vorkehrungen zu deren Verringerung getroffen. Zudem ist zu bedenken, dass diese Meta-Evaluation in erster Linie dem Erkenntnisgewinn diente. Hätte es sich um eine beauftragte Meta-Evaluation gehandelt, die z. B. aufgrund von Unstimmigkeiten erforderlich gewesen wäre, dann wäre womöglich ein noch objektiveres Urteil anzustreben gewesen, wobei der Einbezug externer, am Programm und der Evaluation unbeteiligter Personen sinnvoll gewesen wäre. Im vorliegenden Fall stand jedoch die Erarbeitung von Empfehlungen im Fokus, sodass die Bewertung selbst, also welche Standards von der

Evaluation wie gut eingehalten worden waren, als sekundär zu betrachten ist. Vielmehr ging es um eine möglichst vollständige Abbildung der Stärken und Schwächen der Evaluation. Daher sind die Maßnahmen, die hier zur Erreichung einer möglichst gerechten Beurteilung ergriffen wurden, insgesamt als angemessen zu bewerten. Zur Verifizierung der vorliegenden Bewertung wäre dennoch eine zusätzliche externe Meta-Evaluation wünschenswert und auch im Hinblick auf eine Weiterentwicklung und mögliche Verbesserung der Meta-Evaluationsmethodik interessant. Anhand eines Abgleichs von interner und externer Evaluationsbewertung könnten beispielsweise typische Verzerrungen bei internen Meta-Evaluationen aufgedeckt werden.

Wie bereits angesprochen, können bei Meta-Evaluationen Interessenkonflikte auftreten. Dies gilt sowohl für interne als auch für externe Evaluationsbewertungen (Yarbrough et al., 2011). In der Tat ist eine kritische Auseinandersetzung mit der eigenen Arbeit nicht immer angenehm, kann jedoch die Kompetenzentwicklung von Evaluatoren fördern (Abschnitt 2.1.5). In der vorliegenden Arbeit wurde es als enorm wertvoll empfunden, die Stärken und Schwächen der Evaluation kritisch zu reflektieren und damit die eigenen Kompetenzen weiterzuentwickeln. Dass im Rahmen der Evaluationsbewertung die gemachten Erfahrungen explizit gemacht werden mussten, trug ebenfalls zu diesem Lernprozess bei, der somit in gewissem Ausmaß auch bei den Stellungnehmern stattgefunden haben dürfte. Daher wird in Anlehnung an andere Autoren (DeGEval, 2008b; Spiel et al., 2012) empfohlen, bei jeder Evaluation eine Aufbereitung der gemachten Erfahrungen durch die Evaluierenden und andere Stakeholder vorzunehmen (also obligatorisch zumindest eine kurze Meta-Evaluation vorzunehmen), um ein Lernen aus diesen Erfahrungen zu ermöglichen.

> **Empfehlung 35**: Bei jeder Evaluation sollte seitens der Evaluatoren sowie anderer Stakeholder ein Reflexionsprozess hinsichtlich der Stärken und Schwächen des Projekts stattfinden (im Sinne einer obligatorischen knappen Meta-Evaluation), damit aus den Erfahrungen gelernt werden kann.

Ein Vorteil dieser internen Meta-Evaluation bestand darin, dass als Grundlage der Bewertung umfangreiche Erfahrungen und Unterlagen aus dem gesamten Evaluationsprozess zur Verfügung standen. Diese ermöglichten, gemeinsam mit den Stellungnahmen, eine Einschätzung

der Evaluation auf allen Bewertungsdimensionen. Demgegenüber stehen externen Meta-Evaluatoren teils nur die Evaluationsberichte und ggf. weitere Unterlagen zur Verfügung, die jedoch häufig nicht genügend Informationen enthalten, um eine umfassende Bewertung aller Aspekte zu erlauben (z. B. Bührlen-Armstrong, 1998; Widmer, 1996). Daher wird in der Literatur eine detaillierte Dokumentation von Evaluationen gefordert, um Meta-Evaluationen zu ermöglichen (Abschnitt 2.2.5.1). Diese Forderung soll hiermit bekräftigt werden, da einerseits mehr Meta-Evaluationen anzustreben sind (vgl. Abschnitt 5.3.2), die dadurch unterstützt würden, und andererseits Evaluierende sich durch ein Festhalten ihres Vorgehens auch absichern können. Allerdings ist zu bedenken, dass eine ausführliche Dokumentation aller Arbeitsschritte viel Aufwand für die Evaluierenden bedeutet (vgl. Widmer, 1996). Zudem müsste den Evaluatoren bekannt gemacht werden, welche Punkte es zu dokumentieren gilt. Während in Evaluationsberichten vermutlich vorwiegend methodische Aspekte behandelt werden, wäre unter anderem die Aufzeichnung von datenschutzrechtlichen und faire Interaktionen betreffenden Belangen wichtig, um die entsprechenden Fairnessstandards bewerten zu können. Die Dokumentation des Evaluationsprozesses könnte dabei ebenso von den Evaluationsstandards geleitet werden wie die Evaluation selbst.

Empfehlung 36: Um Meta-Evaluationen zu ermöglichen, sollten Evaluationsstudien im Detail dokumentiert werden. Standards für Evaluationen können diese Dokumentationen strukturieren.

Erwähnt werden sollte weiterhin, dass diese Meta-Evaluation mit einigem Aufwand verbunden war. Dies betraf einerseits die Leistungen der Meta-Evaluatorin, die sich intensiv in die Evaluationsstandards einarbeitete, die gesamte, umfangreiche Evaluation bewertete, die Stellungnahmen erhob und auswertete, diese Informationen sowie die Rückmeldungen aus den zweiten Stellungnahmen in ihre Bewertung integrierte und die Meta-Evaluation möglichst detailliert beschrieb, um ihr Vorgehen und die Bewertung nachvollziehbar zu machen. Andererseits bedeuteten die Stellungnahmen auch für die Stellungnehmer zeitlichen Aufwand, für den sie nicht entschädigt wurden. Zwar ist anzumerken, dass es sich bei der vorliegenden Studie um die systematische Bewertung eines sehr umfangreichen Evaluationsprojekts mit vielen Erhebungen und etlichen Beteiligten handelte und sie deshalb

vermutlich mit mehr Arbeit einherging als dies bei Bewertungen kleinerer Evaluationen der Fall wäre. Jedoch gestaltet sich allein die Einarbeitung in die Evaluationsstandards zeitintensiv (Abschnitt 5.2.1), sodass Meta-Evaluationen, sofern sie einen Mindestanspruch an Systematik erfüllen, einen nicht unerheblichen Aufwand bedeuten. Der Forderung nach mehr Meta-Evaluationen steht somit die Frage nach deren Finanzierung gegenüber: Wer bezahlt diesen Aufwand? Meta-Evaluationen bieten Lernpotential für Evaluatoren und andere Beteiligte und sie dienen dem Ziel, die Evaluationspraxis zu verbessern (Abschnitt 2.2.4). Jedoch erscheint es unrealistisch zu erwarten, dass Auftraggeber von Evaluationen auch Evaluationsbewertungen finanzieren, zumindest solche, die „lediglich" dem Erkenntnisgewinn in evaluationsmethodischer Hinsicht gewidmet sind. Gerade hier wären jedoch mehr Meta-Evaluationen wünschenswert (Abschnitt 5.3.2). Eine Lösung könnte die Durchführung weiterer Meta-Evaluationen in Form von Qualifikationsarbeiten darstellen. In diesem Kontext entstanden nicht nur die vorliegende Arbeit, sondern auch andere umfangreiche Evaluationsbewertungen (Bührlen-Armstrong, 1998; Widmer, 1996). Dem Vorteil der geringen Kosten steht der potentielle Nachteil der Unerfahrenheit der Meta-Evaluatoren gegenüber. Dieser könnte jedoch durch eine kompetente Betreuung durch erfahrene (Meta-)Evaluatoren ausgeglichen werden.

Empfehlung 37: Der Forderung nach weiteren Meta-Evaluationen könnte dadurch nachgekommen werden, dass mehr Evaluationsbewertungen in Form von Qualifikationsarbeiten realisiert werden.

5.4 Fazit zur vorliegenden Studie und Ausblick

In der vorliegenden Arbeit wurde die Evaluation des Bundesprogramms „Elternchance ist Kinderchance" bewertet. Da es sich um eine interne Meta-Evaluation handelte, war eine sehr detaillierte Beurteilung möglich. Die angewendete Methode, bei der auch Stellungnahmen verschiedener Stakeholder eingeholt wurden, kann als aufwendig, aber nützlich eingestuft werden, insbesondere hinsichtlich der Systematik und des Einbezugs mehrerer Perspektiven. Es wurden

einige Maßnahmen zur Reduzierung der Subjektivität getroffen, um eine möglichst faire Bewertung zu erhalten. Da mit der Evaluation von „Elternchance ist Kinderchance“ ein umfangreiches Projekt mit vielen Beteiligten und umfassender Datenerhebung vorlag, konnten im Rahmen der Meta-Evaluation etliche Erfahrungen und Herausforderungen beschrieben und Empfehlungen für Evaluatoren und Auftraggeber generiert werden. Zudem wurden die deutschen Standards für Evaluation als Instrument für Evaluationsbewertungen getestet und wichtige Erkenntnisse zu diesen Richtlinien gewonnen. Auch Empfehlungen für künftige Meta-Evaluationen wurden erarbeitet. Hinsichtlich der Empfehlungen ist zu beachten, dass diese nicht für alle künftigen Evaluationen Gültigkeit besitzen werden, wenngleich bei ihrer Generierung auch andere Studien einbezogen wurden. Schließlich wurde die angewandte Methode möglichst genau dargestellt und kritisch reflektiert, da ein Ziel der Arbeit darin bestand, die Diskussion zur Meta-Evaluations-Methodik anzustoßen.

Die vorliegende Studie erbrachte somit viele bedeutsame Erkenntnisse, die in Empfehlungen für die (Meta-)Evaluationspraxis und Hinweisen zu den Evaluationsstandards resultierten. Neben diesen praktischen Implikationen wurde auch Bedarf an weiterer Forschungsaktivität deutlich. So sind weitere Meta-Evaluationen angezeigt, um die hier ausgesprochenen Ratschläge zu ergänzen, sie zu unterstützen oder evtl. auch zu widerlegen. Dabei sollten auch Evaluationen bewertet werden, die unter anderen Rahmenbedingungen und in anderen Evaluationsgebieten durchgeführt wurden als die in vorliegender Arbeit untersuchte Evaluation. Nur dadurch kann auf lange Sicht ein auf einer breiten Erfahrungsgrundlage basierender Fundus an Ratschlägen generiert werden, der Evaluatoren Orientierung bietet. Diese Ratschläge können die bereits existierenden Evaluationsstandards ergänzen und/oder konkretisieren, evtl. auch in Richtung der mitunter geforderten feldspezifischen Standards. Auch Orientierungshilfen zur Priorisierung der Standards sind denkbar, wenn einigen Richtlinien metaevaluationsübergreifend mehr Gewicht eingeräumt wird als anderen. Dazu bedarf es jedoch der Durchführung von weiteren Meta-Evaluationen (vgl. Stufflebeam, 2001b; Widmer, 1996). Dies könnte auch dazu führen, dass die Methodik von Evaluationsbewertungen weiterentwickelt wird, denn auch hierzu gibt es bislang insbesondere aus dem deutschen Sprachraum nur wenige Publikationen. Ein erster Schritt in diese Richtung wären Meta-Evaluationsberichte, die die angewendete

Methode klarer schildern, als dies bisher der Fall ist. Auch eine kritische Reflexion der jeweiligen Methode, wie sie in vorliegender Arbeit vorgenommen und berichtet wurde, wäre für die Methodenentwicklung förderlich. Als weiterer Vorteil von mehr Evaluationsbewertungen schließlich ist zu benennen, dass sie auch eine Qualitätsverbesserung von Evaluationsstudien bewirken können. Immerhin stellt diese den zentralen Zweck von Meta-Evaluationen dar und vermutlich führt allein das Bewusstsein, dass eine Evaluation bewertet werden könnte, zu einer verbesserten Durchführung derselben. Wenn Evaluierende der Forderung nach genaueren Evaluationsberichten bzw. einer detaillierten Dokumentation des Evaluationsprozesses nachkommen, erleichtern sie damit nicht nur externe Meta-Evaluationen, sondern führen auch wahrscheinlicher qualitativ hochwertige Studien durch (vgl. Wittmann, 1985). Meta-Evaluationen fördern somit in mehrerlei Hinsicht die Qualitätssicherung von Evaluationen.

6 Literaturverzeichnis

Alexander, H. (2003). Health-service evaluations: Should we expect the results to change practice? *Evaluation, 9,* 405-414.

American Evaluation Association. (2004). *Guiding principles for evaluators.* Zugriff am 06.05.2015. Verfügbar unter http://www.eval.org/p/cm/ld/fid=51

Atria, M., Reimann, R. & Spiel, C. (2006). Qualitätssicherung durch Evaluation. Die Bedeutung von Zielexplikation und evaluativer Haltung. In C. Steinebach (Hrsg.), *Handbuch Psychologische Beratung* (S. 574-586). Stuttgart: Klett-Cotta.

Balthasar, A. (2009). Evaluationen in der Schweiz: Verbreitung und Verwendung. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 486-497). Wiesbaden: VS Verlag für Sozialwissenschaften.

Beck, M., Schwaß, M. & Stemmler, M. (2016). *Nachbericht zur Evaluation des Bundesprogramms „Elternchance ist Kinderchance – Elternbegleitung der Bildungsverläufe der Kinder". Bericht zur Kontrollgruppe im Modul 3 – die Wirkung auf Ebene der Eltern und Kinder*. Erlangen: FAU University Press.

Beywl, W. (2001). Konfliktfähigkeit der Evaluation und die 'Standards für Evaluationen'. *Sozialwissenschaften und Berufspraxis, 24* (2), 151-164.

Beywl, W. (2012). Evaluation ist systematische Kritik. Nachvollziehbar bewerten und Wertmaßstäbe offen legen. *Sozial Extra, 36* (9/10), 15-17.

Beywl, W., Kehr, J., Mäder, S. & Niestroj, M. (2007). *Evaluation Schritt für Schritt: Planung von Evaluationen*. Heidelberg: hiba.

Beywl, W. & Speer, S. (2003). *Was kennzeichnet gute Evaluationen in der Berufsbildung? Standards für Evaluation und Perspektiven ihrer Weiterentwicklung.* Zugriff am 11.06.2015. Verfügbar unter http://www.bibb.de/veroeffentlichungen/de/bwp/show/id/1678

Beywl, W. & Speer, S. (2004). Developing standards to evaluate vocational education and training programmes. In P. Descy & M. Tessaring (Hrsg.), *The foundations of evaluation and impact research. Third*

report on vocational training research in Europe: Background report. Luxembourg: Office for Official Publications of the European Communities (Cedefop Reference series, 3040).

Beywl, W. & Widmer, T. (2006). Die ‚Standards' im Vergleich mit weiteren Regelwerken zur Qualität fachlicher Leistungserstellung. In Joint Committee on Standards for Educational Evaluation (JCSEE) (Hrsg.), *Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation"* (3. Aufl., S. 263-299). Wiesbaden: VS Verlag für Sozialwissenschaften.

Bortz, J. & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4. Aufl.). Heidelberg: Springer Medizin Verlag.

Böttcher, W., Kerlen, C., Maats, P., Schwab, O. & Sheikh, S. (2014a). Arbeitsfelder und Herausforderungen der Evaluation. In W. Böttcher, C. Kerlen, P. Maats, O. Schwab & S. Sheikh (Hrsg.), *Evaluation in Deutschland und Österreich. Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation* (S. 7-14). Münster: Waxmann.

Böttcher, W., Kerlen, C., Maats, P., Schwab, O. & Sheikh, S. (Hrsg.). (2014b). *Evaluation in Deutschland und Österreich. Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation*. Münster: Waxmann.

Brandt, T. (2007). Sozialer Kontext der Evaluation. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 164-194). Münster: Waxmann.

Brandt, T. (2009). *Evaluation in Deutschland. Professionalisierungsstand und -perspektiven*. Münster: Waxmann.

Bührlen-Armstrong, B. (1998). *Evaluationsforschung und Qualitätssicherung Evaluationsforschung und Qualitätssicherung in der personalkommunikativen AIDS-Prävention. Eine Meta-Evaluation*. Unveröffentlichte Dissertation, Albert-Ludwigs-Universität. Freiburg i. Br.

Caspari, A. (2010). *Lernen aus Evaluierungen. Meta-Evaluation & Evaluationssynthese von InWEnt-Abschlussevaluierungen 2009*. Zugriff am 26.05.2015. Verfügbar unter http://www.giz.de/de/ueber_die_giz/516.html

Caspari, A. (2011). *Meta-Evaluation & Evaluationssynthese 2011 - Hauptbericht. Querschnittsauswertung dezentraler Evaluierungen der Human Capacity Development Programme des Jahres 2010.* Zugriff am 11.06.2015. Verfügbar unter http://www.giz.de/de/ueber_die_giz/516.html

Cook, T. D. & Gruder, C. L. (1978). Metaevaluation research. *Evaluation Review, 2,* 5-51.

Cooksy, L. J. & Caracelli, V. J. (2005). Quality, context, and use. Issues in achieving the goals of metaevaluation. *American Journal of Evaluation, 26,* 31-42.

Döring, N. (2009). Phasen der Evaluationsforschung. In H. Holling (Hrsg.), *Grundlagen und statistische Methoden der Evaluationsforschung* (Reihe: Enzyklopädie der Psychologie. Methodologie und Methoden. Evaluation, Bd. 1, S. 99-134). Göttingen: Hogrefe.

Erzberger, C. & Kelle, U. (2014). Methodisch handeln im Evaluationsprozess. In W. Böttcher, C. Kerlen, P. Maats, O. Schwab & S. Sheikh (Hrsg.), *Evaluation in Deutschland und Österreich. Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation* (S. 171-177). Münster: Waxmann.

Fabian, C. (2009). Nutzung der Evaluationsfunktion im Vergleich. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 498-505). Wiesbaden: VS Verlag für Sozialwissenschaften.

Finn, C. E., Stevens, F. I., Stufflebeam, D. L. & Walberg, H. J. (1997). A meta-evaluation. *International Journal of Educational Research, 27,* 159-174.

Fitzpatrick, J. L., Sanders, J. R. & Worthen, B. R. (2012). *Program evaluation. Alternative approaches and practical guidelines* (4. Aufl.). Upper Saddle River, NJ: Pearson Education.

Friede, C. K. (1987). Probleme von Schulvergleichsuntersuchungen, ihres Kontextes und Umfeldes. In K. Aurin (Hrsg.), *Schulvergleich in der Diskussion* (1. Aufl., S. 153-170). Stuttgart: Klett-Cotta.

Gerleigner, S., Guglhör-Rudan, A., Hein, K., Müller, D., Walper, S., Beck, M. et al. (2014). *Das Bundesprogramm „Elternchance ist Kinderchance". Zentrale Befunde der Evaluation.* Berlin: BMFSFJ.

Gesellschaft für Evaluation e. V. (n.d.a). *Evaluation und Gesellschaft. Positionspapier der DeGEval - Gesellschaft für Evaluation.* Zugriff am 19.02.2015. Verfügbar unter http://www.degeval.de/publikationen/positionspapiere/

Gesellschaft für Evaluation e. V. (n.d.b). *Steuerung braucht Evaluation. Positionspapier der DeGEval – Gesellschaft für Evaluation.* Zugriff am 16.02.2015. Verfügbar unter http://www.degeval.de/publikationen/positionspapiere/

Gesellschaft für Evaluation e. V. (2002). *Angefragte Gutachten / Meta-Evaluationen auf Basis der DeGEval-Standards*. Vom Vorstand der DeGEval im April 2002 verabschiedetes, unveröffentlichtes Papier.

Gesellschaft für Evaluation e. V. (2004a). *Befragung zur Steuerung des Revisionsprozesses der DeGEval-Standards für Evaluation (REVAL). Bericht zur qualitativen Datenauswertung.* Zugriff am 09.06.2015. Verfügbar unter http://www.degeval.de/degeval-standards/revisionsprozess/

Gesellschaft für Evaluation e. V. (2004b). *Befragung zur Steuerung des Revisionsprozesses der DeGEval-Standards für Evaluation (REVAL). Bericht zur quantitativen Datenauswertung.* Zugriff am 09.06.2015. Verfügbar unter http://www.degeval.de/degeval-standards/revisionsprozess/

Gesellschaft für Evaluation e. V. (2004c). *Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation.* Zugriff am 06.05.2015. Verfügbar unter http://www.degeval.de/publikationen/selbstevaluation/

Gesellschaft für Evaluation e. V. (2007). *Empfehlungen für Auftraggebende von Evaluationen. Eine Einstiegsbroschüre für den Bereich der Öffentlichen Verwaltung.* Zugriff am 16.02.2015. Verfügbar unter http://www.degeval.de/publikationen/auftraggebende-von-evaluationen/

Gesellschaft für Evaluation e. V. (2008a). *Empfehlungen für die Aus- und Weiterbildung in der Evaluation. Anforderungsprofile an Evaluatorinnen und Evaluatoren* (4. Aufl.). Zugriff am 27.02.2015. Verfügbar unter http://www.degeval.de/publikationen/aus-und-weiterbildung/

Gesellschaft für Evaluation e. V. (2008b). *Standards für Evaluation.* Zugriff am 10.02.2015. Verfügbar unter http://www.degeval.de/publikationen/

Gollwitzer, M. & Jäger, R. S. (2014). *Evaluation kompakt* (2. Aufl.). Weinheim: Beltz.

Greene, J. C. (1992). A case study of evaluation auditing as metaevaluation. *Evaluation and Program Planning, 15,* 71-74.

Gutknecht-Gmeiner, M., Hense, J., Kihm, S., Lüth, K. & Steckhan, H. (2011). *Didaktik der Aus- und Weiterbildung in der Evaluation. Ein Empfehlungspapier des Arbeitskreises Aus- und Weiterbildung in der Evaluation in der Gesellschaft für Evaluation – DeGEval e.V.* Zugriff am 27.02.2015. Verfügbar unter http://www.degeval.de/arbeitskreise/aus-und-weiterbildung-in-der-evaluation/publikationen-und-produkte/

Hanssen, C. E., Lawrenz, F. & Dunet, D. O. (2008). Concurrent meta-evaluation: A critique. *American Journal of Evaluation, 29,* 572-582.

Hense, J. U. & Steckhan, H. (2014). Aus- und Weiterbildung in Evaluation: Vergangenheit – Gegenwart – Zukunft? In W. Böttcher, C. Kerlen, P. Maats, O. Schwab & S. Sheikh (Hrsg.), *Evaluation in Deutschland und Österreich. Stand und Entwicklungsperspektiven in den Arbeitsfeldern der DeGEval – Gesellschaft für Evaluation* (S. 191-205). Münster: Waxmann.

Holling, H. (2009). Grundlagen der Evaluationsforschung. In H. Holling (Hrsg.), *Grundlagen und statistische Methoden der Evaluationsforschung* (Reihe: Enzyklopädie der Psychologie. Methodologie und Methoden. Evaluation, Bd. 1, S. 1-34). Göttingen: Hogrefe.

Hornbostel, S. (2010). (Forschungs-)Evaluation. In D. Simon, A. Knie & S. Hornbostel (Hrsg.), *Handbuch Wissenschaftspolitik* (1. Aufl., S. 293-309). Wiesbaden: VS Verlag für Sozialwissenschaften.

Joint Committee on Standards for Educational Evaluation. (2003). *The student evaluation standards. How to improve evaluations of students*. Thousand Oaks, CA: Corwin Press.

Joint Committee on Standards for Educational Evaluation (Hrsg.). (2006). *Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation"* (3. Aufl.). Wiesbaden: VS Verlag für Sozialwissenschaften.

Joint Committee on Standards for Educational Evaluation. (2009). *The personnel evaluation standards. How to assess systems for evaluating educators* (2. Aufl.). Thousand Oaks, CA: Corwin Press.

Keller-Ebert, C. (2006). Ziehen Auftraggebende Nutzen aus Evaluationen? Ergebnisse einer Follow-up-Studie. *Sozialwissenschaften und Berufspraxis, 29* (1), 101-107.

Kuhn, J., Lampert, T. & Ziese, T. (2012). Einführung ins Thema: Komplexe Interventionen – komplexe Evaluationen? In Robert Koch-Institut, Bayerisches Landesamt für Gesundheit und Lebensmittelsicherheit (Hrsg.), *Evaluation komplexer Interventionsprogramme in der Prävention: Lernende Systeme, lehrreiche Systeme? Beiträge zur Gesundheitsberichterstattung des Bundes* (S. 9-14). Berlin: Robert Koch-Institut.

Kury, H. & Obergfell-Fuchs, J. (2012). Evaluationsforschung. In G. Albrecht & A. Groenemeyer (Hrsg.), *Handbuch soziale Probleme* (2. Aufl., S. 1521-1555). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lösel, F., Köferl, P. & Weber, F. (1987). *Meta-Evaluation in der Sozialtherapie. Qualitative und quantitative Analysen zur Behandlungsforschung in sozialtherapeutischen Anstalten des Justizvollzugs*. Stuttgart: Ferdinand Enke Verlag.

Lösel, F., Schmucker, M., Plankensteiner, B. & Weiss, M. (2006). *Bestandsaufnahme und Evaluation von Angeboten im Elternbildungsbereich*. Berlin: Bundesministerium für Familie, Senioren, Frauen und Jugend.

Love, A. & Russon, C. (2005). Evaluation standards in an international context. In C. Russon & G. Russon (Hrsg.), *International perspectives on evaluation standards* (New directions for evaluation, no. 104, S. 5-14). San Francisco: Jossey-Bass.

Lynch, D. C., Greer, A. G., Larson, L. C., Cummings, D. M., Harriett, B. S., Springer Dreyfus, K. et al. (2003). Descriptive metaevaluation: Case study of an interdisciplinary curriculum. *Evaluation and The Health Professions, 26,* 447-461.

Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11. Aufl.). Weinheim: Beltz.

Meyer, W. (2002). *Was ist Evaluation?* Zugriff am 29.01.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=189

Meyer, W. (2007). Evaluationsdesigns. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 143-163). Münster: Waxmann.

Mittag, W. (2003). Qualitätssicherung und Evaluation. In M. Jerusalem & H. Weber (Hrsg.), *Psychologische Gesundheitsförderung – Diagnostik und Prävention* (1. Aufl., S. 121-137). Göttingen: Hogrefe.

Mittag, W. & Hager, W. (2000). Ein Rahmenkonzept zur Evaluation psychologischer Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien* (1. Aufl., S. 102-128). Bern: Verlag Hans Huber.

Moosbrugger, H. & Schweizer, K. (2002). Evaluationsforschung in der Psychologie. *Zeitschrift für Evaluation, 1,* 19-37.

Müller, D., Beck, M., Gerleigner, S., Guglhör-Rudan, A., Hein, K., Schwaß, M. et al. (2015). *Evaluation des Bundesprogramms „Elternchance ist Kinderchance - Elternbegleitung der Bildungsverläufe der Kinder". Abschlussbericht.* München: DJI.

Newman, D. L. & Brown, R. D. (1992). Violations of evaluation standards: Frequency and seriousness of occurrence. *Evaluation Review, 16,* 219-234.

Patry, J.-L. & Hager, W. (2000). Abschließende Bemerkungen: Dilemmata in der Evaluation. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien* (1. Aufl., S. 258-275). Bern: Verlag Hans Huber.

Raven, J. (1984). Some limitations of the standards. *Evaluation and Program Planning, 7,* 363-370.

Rossi, P. H., Freeman, H. E. & Hofmann, G. (1988). *Programm-Evaluation. Einführung in die Methoden angewandter Sozialforschung*. Stuttgart: Ferdinand Enke Verlag.

Rost, J. (2000). Allgemeine Standards für die Evaluationsforschung. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien* (1. Aufl., S. 129-140). Bern: Verlag Hans Huber.

Roth, R. (2012). Sind kritische Evaluationen möglich? Skeptische Anmerkungen aus der Evaluationspraxis. *Sozial Extra, 36* (7-8), 16-19.

Ruhe, V. & Boudreau, J. D. (2013). The 2011 Program Evaluation Standards: A framework for quality in medical education programme evaluations. *Journal of Evaluation in Clinical Practice, 19,* 925-932.

Russon, C. & Russon, G. (Hrsg.). (2005). *International perspectives on evaluation standards* (New directions for evaluation, no. 104). San Francisco: Jossey-Bass.

Schendera, C. F. G., Janz, F., Klauß, T. & Lamers, W. (2003). Die Umsetzung von Evaluationskriterien im Projekt "Perspektiven der schulischen Bildungs- und Erziehungsrealität von Kindern und Jugendlichen mit schweren und mehrfachen Behinderungen" (BiSB-Projekt). *Zeitschrift für Evaluation, 2,* 223-232.

Schiffler, A. & Hübner, S. (2000). Allgemeine Standards für die Evaluationspraxis. Die Standards des „Joint Committee on Standards for Educational Evaluation“ und ihre Anwendung auf praktische Aspekte bei der Evaluation von psychologischen Interventionsmaßnahmen. In W. Hager, J.-L. Patry & H. Brezing (Hrsg.), *Handbuch Evaluation psychologischer Interventionsmaßnahmen. Standards und Kriterien* (1. Aufl., S. 141-152). Bern: Verlag Hans Huber.

Schwandt, T. A. & Halpern, E. S. (1988). *Linking auditing and metaevaluation. Enhancing quality in applied research* (Applied social research methods series, vol. 11). Newbury Park, CA: Sage.

Scott-Little, C., Hamann, M. S. & Jurs, S. G. (2002). Evaluations of afterschool programs: A meta-evaluation of methodologies and narrative synthesis of findings. *American Journal of Evaluation, 23,* 387-419.

Spiel, C. & Bergsmann, E. (2009). Zur Nutzung der Evaluationsfunktion: Partielle Bestandsaufnahme und Pilotstudie in österreichischen und deutschen Bundesministerien. In T. Widmer, W. Beywl & C. Fabian (Hrsg.), *Evaluation. Ein systematisches Handbuch* (S. 477-485). Wiesbaden: VS Verlag für Sozialwissenschaften.

Spiel, C. & Gössler, M. (2001). Zwischen Selbstzweck und Qualitätsmanagement – Quo vadis, evaluatione? In C. Spiel (Hrsg.), *Evaluation universitärer Lehre – zwischen Qualitätsmanagement und Selbstzweck* (S. 9-20). Münster: Waxmann.

Spiel, C., Gradinger, P. & Lüftenegger, M. (2010). Grundlagen der Evaluationsforschung. In H. Holling & B. Schmitz (Hrsg.), *Handbuch Statistik, Methoden und Evaluation* (S. 223-232). Göttingen: Hogrefe.

Spiel, C., Popper, V. & Kollmayer, M. (2012). *Leitfaden für die Evaluation komplexer Verbundprojekte. Leitfaden im Auftrag der Robert Bosch Stiftung*.

Stamm, M. (2003). Evaluation im Spiegel ihrer Nutzung: Grande idée oder grande illusion des 21. Jahrhunderts? *Zeitschrift für Evaluation, 2,* 183-200.

Stockmann, R. (2002a). *Evaluation als integriertes Lehr- und Forschungsprogramm.* Zugriff am 30.01.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=189

Stockmann, R. (2002b). *Selbstdarstellung des CEVAL.* Zugriff am 30.01.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=189

Stockmann, R. (2004). *Was ist eine gute Evaluation? Einführung zu Funktionen und Methoden von Evaluationsverfahren.* Zugriff am 12.02.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=189

Stockmann, R. (2007). Einführung in die Evaluation. In R. Stockmann (Hrsg.), *Handbuch zur Evaluation. Eine praktische Handlungsanleitung* (S. 24-70). Münster: Waxmann.

Stockmann, R. (2010). *10 Thesen zur Evaluation der Wirksamkeit der Entwicklungszusammenarbeit.* Zugriff am 16.02.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=189

Stockmann, R. (2011). *Evaluation – eine Begriffsdefinition.* Zugriff am 04.02.2015. Verfügbar unter http://www.ceval.de/typo3/index.php?id=24

Stufflebeam, D. L. (1978). Meta evaluation: An overview. *Evaluation and The Health Professions, 1* (1), 17-43.

Stufflebeam, D. L. (1981). Metaevaluation: Concepts, standards, and uses. In R. A. Berk (Hrsg.), *Educational evaluation methodology: The state of the art* (S. 146-163). Baltimore, MD: The Johns Hopkins University Press.

Stufflebeam, D. L. (1999a). *Program evaluations metaevaluation checklist. Based on The Program Evaluation Standards.* Zugriff am 06.05.2015. Verfügbar unter http://www.wmich.edu/evalctr/checklists/evaluation-checklists/

Stufflebeam, D. L. (1999b). Using professional standards to legally and ethically release evaluation findings. *Studies in Educational Evaluation, 25,* 325-334.

Stufflebeam, D. L. (2000). The methodology of metaevaluation as reflected in metaevaluations by the Western Michigan University Evaluation Center. *Journal of Personnel Evaluation in Education, 14,* 95-125.

Stufflebeam, D. L. (2001a). Evaluation checklists: Practical tools for guiding and judging evaluations. *American Journal of Evaluation, 22,* 71-79.

Stufflebeam, D. L. (2001b). The metaevaluation imperative. *American Journal of Evaluation, 22,* 183-209.

Stufflebeam, D. L. (2004). A note on the purposes, development, and applicability of the Joint Committee Evaluation Standards. *American Journal of Evaluation, 25,* 99-102.

Stufflebeam, D. L., Goodyear, L., Marquart, J. & Johnson, E. (2005). *Guiding principles checklist for evaluating evaluations. Based on the 2004 guiding principles for evaluators.* Zugriff am 06.05.2015. Verfügbar unter http://www.wmich.edu/evalctr/checklists/evaluation-checklists/

Ulrich, S. & Wenzel, F. M. (2003). *Partizipative Evaluation. Ein Konzept für die politische Bildung*. Gütersloh: Verlag Bertelsmann-Stiftung.

Walper, S., Jurczyk, K., Bradna, M., Guglhör-Rudan, A., Koenigsbeck, D., Müller, D. et al. (2013). *Evaluation des Bundesprogramms „Elternchance ist Kinderchance. Elternbegleitung der Bildungsverläufe der Kinder". Zwischenbericht,* München und Erlangen.

Weiss, M., Schmucker, M. & Lösel, F. (2015). Meta-Analyse zur Wirkung familienbezogener Präventionsmaßnahmen in Deutschland. *Zeitschrift für Klinische Psychologie und Psychotherapie, 44* (1), 27-44.

Widmer, T. (1996). *Meta-Evaluation. Kriterien zur Bewertung von Evaluationen*. Bern: Haupt.

Widmer, T. (2005). The development and status of evaluation standards in Western Europe. In C. Russon & G. Russon (Hrsg.), *International perspectives on evaluation standards* (New directions for evaluation, no. 104, S. 31-42). San Francisco: Jossey-Bass.

Widmer, T. (2006). Evaluationsstandards in Europa. In Joint Committee on Standards for Educational Evaluation (JCSEE) (Hrsg.), *Handbuch der Evaluationsstandards. Die Standards des "Joint Committee on Standards for Educational Evaluation"* (3. Aufl., S. 301-312). Wiesbaden: VS Verlag für Sozialwissenschaften.

Widmer, T. (2008). Evaluationsansätze und ihre Effekte: Erfahrungen aus verschiedenen Politikfeldern. In H. Matthies & D. Simon (Hrsg.), *Wissenschaft unter Beobachtung. Effekte und Defekte von Evaluationen* (1. Aufl., S. 267-287). Wiesbaden: VS Verlag für Sozialwissenschaften.

Widmer, T., Landert, C. & Bachmann, N. (2000). *Evaluations-Standards der Schweizerischen Evaluationsgesellschaft (SEVAL-Standards).* Zugriff am 30.04.2015. Verfügbar unter http://www.seval.ch/de/standards/index.cfm

Wittmann, W. W. (1985). *Evaluationsforschung. Aufgaben, Probleme und Anwendungen*. Berlin: Springer.

Wottawa, H. & Thierau, H. (1998). *Lehrbuch Evaluation* (2. Aufl.). Bern: Verlag Hans Huber.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K. & Caruthers, F. A. (2011). *The Program Evaluation Standards: A guide for evaluators and evaluation users* (3. Aufl.). Thousand Oaks, CA: Sage.

Znoj, H. & Regli, D. (2006). Begriff und Arten der Evaluation. In B. Renneberg & P. Hammelstein (Hrsg.), *Gesundheitspsychologie* (S. 283-289). Heidelberg: Springer Medizin Verlag.

7 Tabellen- und Abbildungsverzeichnis

Tabellen

Tabelle 1: Aufgaben im Verlauf einer Meta-Evaluation 56
Tabelle 2: Stichprobenumfänge und Ausschöpfungsquoten der drei Module..... 89

Abbildungen

Abbildung 1: Evaluationsarten zu verschiedenen Zeitpunkten im Programmablauf 13
Abbildung 2: Evaluationsphasen und -aufgaben..... 17
Abbildung 3: Anforderungen an die Kompetenzen von Evaluatoren..... 36
Abbildung 4: Evaluationsdesign 87
Abbildung 5: Drop-Out-Raten sowie -Gründe im Rahmen der Rekrutierung der Elternstichprobe durch die Fachkräfte (Modul 3, Interventionsgruppe) 95
Abbildung 6: Drop-Out-Raten sowie -Gründe im Rahmen der Rekrutierung der Elternstichprobe (Modul 3, Interventionsgruppe) 96
Abbildung 7: Übersicht über die Ergebnisse der Meta-Evaluation..... 199

8 Anhang

Anhang: Fragenkatalog für die Stellungnahmen im Rahmen der Meta-Evaluation

Inhaltsbereiche	Fragen	Befragte Stellungnehmer
Planung und Durchführung der Datenerhebung	Wie haben Sie unseren Online-Fragebogen/unser telefonisches Interview wahrgenommen? Bitte gehen Sie dabei auf Dauer, Verständlichkeit und Nachvollziehbarkeit ein.	EB, Eltern
	Wie gut haben Sie sich bei der Planung der Erhebungsinstrumente beteiligt gefühlt? Wären Sie gerne mehr in die Planung der Erhebungsinstrumente einbezogen worden? Bitte begründen Sie Ihre Antworten.	BMFSFJ, Servicestelle, DJI[a], Träger, KL, EB
	Im Rahmen unserer Evaluation haben wir auch Eltern befragt. Wären Sie in die Entwicklung dieser Fragen (des Elternfragebogens) gerne mit einbezogen worden, bzw. fänden Sie dies sinnvoll? Inwiefern?	EB, KL[b]
	Wie bekannt sind Ihnen unsere Erhebungsinstrumente?/Sie kennen auch unsere Erhebungsinstrumente. Haben Sie zu diesen eine Rückmeldung (z. B. hinsichtlich Vollständigkeit, Verständlichkeit, Durchführbarkeit)?	BMFSFJ, Servicestelle, Träger, KL

Fortsetzung

Inhaltsbereiche	Fragen	Befragte Stellungnehmer
Ergebnisse und Nutzen	Fühlen Sie sich über die Ergebnisse der Evaluation angemessen informiert? Haben Sie sich im Verlauf der Evaluation über Zwischenergebnisse angemessen informiert gefühlt? Hätten Sie sich mehr Informationen gewünscht bzw. wünschen Sie sich dies noch für die Zukunft (wenn die endgültigen Ergebnisse feststehen)? In welcher Form?	Alle[a]
	Wie beurteilen Sie die Ergebnisbroschüre „Das Bundesprogramm ‚Elternchance ist Kinderchance' – Zentrale Befunde der Evaluation" hinsichtlich Verständlichkeit, Inhalten und Berücksichtigung unterschiedlicher Sichtweisen?	BMFSFJ, Servicestelle, Träger, KL, EB, Eltern
	Welchen Nutzen ziehen Sie persönlich aus den Ergebnissen der Evaluation? Wie nutzen Sie die Ergebnisse bereits für Ihre Arbeit bzw. wie werden Sie sie künftig noch nutzen? Inwiefern würden Sie sich mehr Unterstützung bei der Umsetzung dieser Ergebnisse in die Praxis wünschen?	Alle[a]
	Welche Nutzung der Evaluationsergebnisse war zu Beginn der Evaluation geplant (z. B. Optimierung des Programms, Entscheidung über die Fortführung des Programms)? Inwieweit konnten diese Erwartungen erfüllt werden und worin lagen ggf. Kompromisse begründet?	BMFSFJ
	Wie empfinden Sie den Nutzen der Evaluation im Verhältnis zu ihren Kosten?	BMFSFJ, DJI
	Welche Erwartungen hatten Sie an die Evaluation? Inwieweit wurden diese erfüllt?[c]	Servicestelle, Träger, KL, EB, Eltern

Fortsetzung

Inhaltsbereiche	Fragen	Befragte Stellungnehmer
Beteiligung am Evaluationsprozess insgesamt	Inwiefern hatten Sie insgesamt das Gefühl, angemessen an der Evaluation beteiligt gewesen zu sein? Bitte begründen Sie Ihre Antwort.	BMFSFJ, Servicestelle, Träger, KL, EB, Eltern
	Wie hätten wir Sie besser am Evaluationsprozess beteiligen können?	BMFSFJ, Servicestelle, Träger, KL, EB, Eltern
Zusammenarbeit mit dem Evaluationsteam	Wie glaubwürdig und kompetent fanden Sie das Evaluationsteam?	Alle[a]
	Wie haben Sie die Zusammenarbeit mit dem Evaluationsteam erlebt?	Alle[a]
Generelle Bewertung und Vorschläge	Gibt es etwas (Negatives/Positives), was noch nicht erwähnt wurde und Sie ergänzen möchten?	Alle
	Was sollte aus Ihrer Sicht bei einer weiteren Evaluation anders gemacht werden?	Alle

Anmerkungen. EB = Elternbegleiter; KL = Kursleiter; Alle = BMFSFJ, Servicestelle, Träger, KL, EB, Eltern, DJI.

[a] Der Vertreter des DJI wurde an diesen Stellen um Einschätzung gebeten, wie gut die Beteiligung der Stakeholder, die Ergebnisvermittlung und -nutzung, die Zusammenarbeit im Evaluationsteam und die Kooperation mit den Beteiligten aus seiner Sicht gelungen war, und wie glaubwürdig die Evaluatoren wahrgenommen wurden.

[b] Die Kursleiter sollten an dieser Stelle angeben, ob sie die Beteiligung von Elternbegleitern beim Elternfragebogen für sinnvoll hielten.

[c] Diese Frage wurde nachträglich ergänzt, als das BMFSFJ seine Stellungnahme bereits abgegeben hatte und konnte diesem daher nicht gestellt werden.

Evaluationsstandards und Meta-Evaluationen stellen wichtige Maßnahmen zur Qualitätssicherung von Evaluationen dar. Allerdings ist insbesondere im deutschen Sprachraum ein diesbezügliches Forschungsdefizit zu verzeichnen. Die vorliegende Studie leistet einen Beitrag zur Schließung dieser Forschungslücke, indem die Evaluation des Bundesprogramms „Elternchance ist Kinderchance“ anhand der deutschen Standards für Evaluation evaluiert wird. Das zentrale Ziel dieser Meta-Evaluation besteht darin, die Herausforderungen im Evaluationsprozess darzulegen, aus ihnen Empfehlungen zu generieren und so die gemachten Erfahrungen für künftige Studien nutzbar zu machen. Darüber hinaus sollen die deutschen Evaluationsstandards unter anderem auf ihre Eignung für Evaluationsbewertungen geprüft sowie Beiträge zur fachlichen Diskussion im Bereich Meta-Evaluationen und Evaluationsstandards geliefert werden. Wie die Meta-Evaluation zeigt, stellt die Evaluation von „Elternchance ist Kinderchance“ ein Beispiel für eine gelungene Evaluationsstudie dar, anhand welcher vielfältige Herausforderungen im Evaluationsverlauf aufgezeigt werden können, die in zahlreichen Empfehlungen für die Evaluationspraxis resultieren. Weiterhin wird die Anwendbarkeit der Evaluationsstandards erörtert und die Richtlinien werden, ebenso wie die angewandte Methode, kritisch beleuchtet und diskutiert.

Das vorliegende Buch richtet sich an Studierende und Forschende, die sich mit dem Thema Evaluation befassen und einen Leitfaden für die Anwendung von Evaluationsstandards bei (Meta-)Evaluationen suchen.

ISBN 978-3-944057-67-5

9 783944 057675

FAU UNIVERSITY PRESS 2016