

Karin Wirth
(Universität Hamburg)

Evaluation neuer Bildungsgänge - Maßnahmen zur Qualitätssicherung zwischen Gegenstandsbereich, Anspruchsgruppen, Prozessverlauf und Innovation

Online unter:

http://www.bwpat.de/ausgabe21/wirth_bwpat21.pdf

in

bwp@ Ausgabe Nr. 21 | Dezember 2011

Qualität und Qualitätsmanagement in der Berufsbildung

Hrsg. von Karin Büchter, Franz Gramlinger & Karl Wilbers
<http://www.bwpat.de> | ISSN 1618-8543

www.bwpat.de

bwp@

Herausgeber von *bwp@* : Karin Büchter, Franz Gramlinger, H.-Hugo Kremer, Ralf Tenberg und Tade Tramm

Berufs- und Wirtschaftspädagogik - online

ABSTRACT (WIRTH 2011 in Ausgabe 21 von *bwp@*)

Online: www.bwpat.de/ausgabe21/wirth_bwpat21.pdf

Bei der Einführung innovativer Bildungsgänge wird berufsbildungspolitisch häufig davon ausgegangen, dass eine wissenschaftlich begleitende Evaluation zur Einhaltung erhöhter Qualitätsansprüche beiträgt. Diese Begleitprogramme stehen dann vor der Herausforderung, für den jeweiligen Gegenstandsbereich Qualitätskriterien entwickeln zu müssen und gleichzeitig den verschiedenen Anspruchsgruppen im Prozessablauf gerecht zu werden. Aus einer übergeordneten Perspektive ist weiterhin zu formulieren, welche Maßstäbe als Qualitätskriterium für Innovationen anzulegen sind. Hinzu kommt, dass bei einer formativ angelegten Evaluation die Grenzen zwischen zu evaluierenden Bildungsgängen und formativen Eingriffen zu klären sind. Schließlich werden die so skizzierten Ebenen in der Auseinandersetzung um Qualitätsstandards häufig nicht deutlich genug voneinander getrennt.

Der Beitrag fokussiert die Frage adäquater Qualitätsmaßstäbe und –maßnahmen anhand eines laufenden Hamburger Schulversuchs, bei dem das Konzept der vollqualifizierenden Berufsfachschule mit der im Bildungssektor anerkannten Erlangung der Fachhochschulreife und einem auf dem Arbeitsmarkt anerkannten dualen Ausbildungsabschluss verknüpft wird. Dazu wird unter Rückgriff auf Überlegungen zum Design-Based Research das Innovationpotenzial dieses neuen Bildungsgangs diskutiert und die Herausforderungen einer formativ und summativ angelegten Evaluation im Hinblick auf die oben skizzierten Ebenen herausgearbeitet.

Evaluation of new education courses – measures for quality assurance between the subject domain, the stakeholder groups, the course of the process and innovation

When innovative education courses are introduced it is often assumed in vocational education policy terms that an academic accompanying evaluation contributes to the maintenance of increased quality demands. These accompanying programmes face the challenge of having to develop quality criteria for the respective subject domain and, at the same time, to meet the requirements of the various stakeholder groups in the course of the process. From an overarching perspective there is also the need to formulate which measurements should be created as quality criteria for innovations. In addition to this, with a formatively created evaluation it is necessary to clarify the parameters between the education courses that are to be evaluated and formative interventions. Finally, these brief outlines of the levels in the debate around quality standards are often not separated from each other clearly enough.

The paper focuses on the question of appropriate quality scales and measures using an ongoing school project in Hamburg. In this project, the concept of the vocational upper school which offers full qualifications is linked with the recognised attainment in the education sector of the qualification for attending technical college, and a dual training qualification which is recognised on the labour market. With recourse to considerations on design based research the innovation potential of this new education course is discussed, and the challenges of a formatively and summatively created evaluation with regard to the levels outlined above are developed.

Evaluation neuer Bildungsgänge - Maßnahmen zur Qualitätssicherung zwischen Gegenstandsbereich, Anspruchsgruppen, Prozessverlauf und Innovation

1 Einführung

Evaluationsforschung ist eine "Variante praxisbezogener Forschung" (LÜDERS 2006, 36), die als besonders „nutzungsorientiert“ (LÜDERS 2006, 39) gilt. Sie ist dabei auf „zielorientiertes Informationsmanagement“ ausgerichtet (KROMREY 2003, 11), das „jede methodisch kontrollierte, verwertungs- und bewertungsorientierte Form des Sammelns, Auswertens und Verwertens von Informationen“ (KROMREY 2003, 11) umfasst. Hauptsächliche Tätigkeiten bei einer Evaluation finden also im Zusammenhang mit dem Bewerten von Maßnahmen, Produkten oder Projekten statt. Wissenschaftliche Evaluationen kennzeichnet laut STOCKMANN (2007, 27), dass "die Bewertung anhand explizit auf den zu evaluierenden Sachverhalt und anhand präzise festgelegter und offengelegter Kriterien vorgenommen wird". Diese Feststellung gilt umso mehr, wenn der Gegenstand der Evaluation in der bildungspolitischen Diskussion kontrovers diskutiert wird. Qualität wird dabei durch Kriterien auf verschiedenen Ebenen gesichert: So müssen einerseits die Datenerhebung und –auswertung bestimmten Qualitätskriterien genügen, andererseits aber auch der Vorgang des Interpretierens und Eingreifens in den zu evaluierenden Bereich zielorientiert verfolgt und offengelegt werden.

Werden die Zielvorgaben für die Evaluation neuer Bildungsgänge genauer analysiert, so lässt sich feststellen, dass mit der Beauftragung einer wissenschaftlichen Begleitung häufig ein doppelter Anspruch verbunden ist. Zum Einen wird erwartet, dass die Evaluation den Erfolg des Bildungsgangs beobachtet, misst und bewertet. Damit wird auch die Hoffnung verbunden, dass allein durch die beobachtende Anwesenheit von Evaluatoren Qualitätsansprüche sichergestellt werden. Zum Anderen ist mit der Evaluation neuer Bildungsgänge nicht selten die Aufgabe verbunden, aktiv in den Implementationsprozess einzugreifen und beratend tätig zu werden. Evaluation kann in diesem Sinne als „Motor der Innovation“ (WILL/WINTELER/ KRAPP 1987, 27) betrachtet werden, in der die Evaluatoren die Rolle von Entwicklern übernehmen.

Dieser an Evaluationen im Bildungsbereich häufig herangetragene doppelte Anspruch von Bewertung und Entwicklung erfordert eine Kombination aus formativer und summativer Evaluation sowie aus Konstruktions- und Implementationsaufgaben, bei denen die Grenzen zwischen zu evaluierendem Gegenstandsbereich und Interventionen präzise auszuloten sind. Unter diesem Aspekt müssen die in der Evaluationsforschung geltenden Qualitätsmaßstäbe um Kriterien für komplexe Konstruktionsaufgaben erweitert werden.

Im folgenden Kapitel wird daher zunächst ein Überblick über bestehende Gütekriterien in der Forschung gegeben, die um allgemein anerkannte Standards für die Evaluation ergänzt wer-

den. Schließlich werden diese Qualitätsmaßstäbe durch Vorgaben des Design-Based Research erweitert.

Im weiteren Verlauf des Beitrags sollen im Kapitel vier anhand der Evaluation eines neuen Bildungsgangs in Hamburg ausgewählte Schwerpunkte angesprochen werden, die als kritische Faktoren einer Qualitätssicherung angesehen werden. Im Kapitel fünf folgt neben einer Übersicht der wichtigsten Qualitätskriterien auf den unterschiedlichen Wirkungsebenen der Evaluation neuer Bildungsgänge ein Fazit in Form von fünf Thesen.

2 Qualitätskriterien für Evaluationen

Weiterreichende Kriterien für die Qualität einer Evaluation finden sich auf unterschiedlichen Ebenen und mit unterschiedlichen Wirkungsgraden. Sie reichen von den „klassischen“ test-theoretischen Gütekriterien (Validität, Reliabilität und Objektivität) über die ergänzenden Kriterien qualitativer Forschung (z.B. intersubjektive Nachvollziehbarkeit, Kohärenz und Relevanz) und Standards der Evaluation (z.B. die Qualitätsstandards der Deutschen Gesellschaft für Evaluation) bis hin zu Design-Kriterien (z.B. Nutzenorientierung, Theoriegeleitetheit, Iterativität). Dabei erhalten nicht alle Kriterien bei der Evaluation neuer Bildungsgänge die gleiche Aussagekraft.

BORTZ und DÖRING (2005, 54) unterscheiden für ein Untersuchungsdesign explorative von hypothesenüberprüfenden Untersuchungen. Den explorativen Untersuchungen ordnen sie u.a. Methoden der offenen Befragung, der Feldbeobachtung und der nonreaktiven Messungen zu. Hypothesenprüfende Untersuchungen hingegen benötigen eine genügend große Stichprobe. Anschließend können Tests durchgeführt und die Hypothesen bestätigt oder verworfen werden. Hiermit ist vorrangig die Aufgabe von Evaluation angesprochen, bei der nach wissenschaftlichen Maßstäben, d.h. unter Bezug auf empirische Daten und unter Verwendung „systematisch vergleichender Verfahren für die Informationsbewertung“ (STOCKMANN 2007, 27) Aussagen über den Bildungsgang zu treffen sind.

Ergänzend zu den quantitativen Gütekriterien definieren qualitativ orientierte Forscher Kernkriterien. In Anlehnung zur intersubjektiven Überprüfbarkeit quantitativer Forscher formuliert z.B. STEINKE (2009, 324) das Kriterium der intersubjektiven Nachvollziehbarkeit, nach dem die Dokumentation des Forschungsverständnisses, der Untersuchungen und der Interpretationen einen erhöhten Stellenwert erhält. Neben anderen Kriterien sollte qualitativ angelegte Forschung kohärent sein, d.h. es sollten Widersprüche in den Daten und Interpretationen bearbeitet werden und die Fragestellung sollte relevant sein, d.h. zur Lösung der Fragestellung beitragen (STEINKE 2009, 330).

Die Evaluation neuer Bildungsgänge unterscheidet sich allerdings in ihrer Auftrags- und Verwertungsorientierung von quantitativer und qualitativer Forschung (vgl. die Diskussion in LÜDERS 2006), so dass dafür weitere Qualitätsmerkmale herangezogen werden müssen. Die DEUTSCHE GESELLSCHAFT FÜR EVALUATION (DeGEval 2011) unterscheidet folgende Standards zur Evaluation: Nützlichkeit bezeichnet die Forderung, dass sich die Eva-

uation an den zu evaluierenden Zwecken und den benötigten Informationen ausrichtet. Die Durchführbarkeit bezieht sich darauf, dass die Evaluation die Ergebnisse angemessen und effizient hervorbringt. Ein Fairnessstandard soll sicherstellen, dass „in einer Evaluation respektvoll und fair mit den betroffenen Personen und Gruppen umgegangen wird“ (DeGEval 2011). Schließlich zielt das Kriterium Genauigkeit auf eine Gültigkeit der Informationen und damit eher auf die interne Validität (BORTZ/ DÖRING 2005, 57) ab.

Wird eine Evaluation mit dem Anspruch durchgeführt, in die bestehende Bildungs- und Unterrichtspraxis einzugreifen und sie zu verändern, dann gilt es zusätzliche Qualitätskriterien zu berücksichtigen, die einerseits auf das Ergebnis, also die Qualität der Unterrichtspraxis abstellen und andererseits den Prozess der Konstruktion und Erstellung in den Blick nehmen. Dazu bieten sich Ansätze an, die derzeit unter dem Begriff des Design-Based Research zu finden sind (vgl. z.B. SHAVELSON et al. 2003). Diese Ansätze entstanden mit dem Anspruch, Lehr-Lernforschung nicht im Labor, sondern in realen Situationen zu untersuchen (vgl. REINMANN 2005, 60). Design-Based Research grenzt sich von Evaluationsforschung dahingehend ab, dass es neben der Lösung eines in der Praxis vorhandenen Problems immer auch um ein „größeres Verständnis von Lernen und Lehren und um die Entwicklung von Theorien“ geht (REINMANN 2005, 63). Ansätze des Design-Based Research diskutieren die Qualität der Eingriffe unter den Fragestellungen, in wie weit sie einen kontinuierlichen Beitrag zur Lösung von in der Praxis vorliegenden Problemen leisten. Als Ergebnis sollen „sowohl kontextualisierte Theorien des Lernens und Lehrens einschließlich Wissen zum Designprozess (theoretischer Output) als auch konkrete Verbesserungen für die Praxis und die Entfaltung innovativer Potenziale im Bildungsalltag (praktischer Output)“ entstehen (REINMANN 2005, 61). Als Qualitätsmaßstäbe gelten dementsprechend neben einem theoriegeleiteten Vorgehen insbesondere die Nutzenorientierung. Bei aller Kritik an diesen Ansätzen (vgl. FISCHER/ WAIBEL/ WECKER 2005, 435f.) können sie dennoch herangezogen werden, wenn es darum geht, Qualitätsmaßstäbe für denjenigen Teil der Evaluation zu formulieren, der als Beratungsaufgabe oder Konstruktionsteil häufig genug forschungsmethodisch unterbewertet wird. In diesem Sinne bieten Ansätze des Design-Based Research mögliche Antworten darauf, wie im Wechsel von Evaluation, Implementation und Konstruktion vorzugehen ist (APREA 2009).

Diese Qualitätsmerkmale werden nun im nächsten Kapitel herangezogen, um an ihnen am Beispiel des Schulversuchs EARA die Herausforderungen deutlich zu machen, mit denen die Evaluation neuer Bildungsgänge im Hinblick auf Gegenstandsbereich, Anspruchsgruppen, Prozessablauf und Innovationsgehalt umzugehen hat. Dabei wird ein Schwerpunkt dieses Beitrags auf dem Gegenstandsbereich liegen, da sich daran zeigen lässt, auf welchen Ebenen Qualitätskriterien wirken müssen, um für eine Evaluation mit dem doppelten Anspruch der Beurteilung und Beratung gelten zu können.

3 Evaluation und Begleitung neuer Bildungsgänge am Beispiel EARA

3.1 Der Schulversuch EARA

Um die Chancen marktbenachteiligter Jugendlicher auf dem Arbeitsmarkt zu erhöhen, wurde in Hamburg ein Schulversuch mit dem Titel „Erprobung neu strukturierter Ausbildungsformen im Rahmen des Ausbildungskonsenses 2007 – 2010“ (EARA) konzipiert sowie in einem technischen und einem kaufmännischen Ausbildungsgang erprobt. Ziel dieses Schulversuchs ist es weiterhin, das Übergangssystem von der Schule in die Ausbildung zu vereinfachen und unproduktive Warteschleifen von Jugendlichen zu vermeiden.

Im Schulversuch EARA absolvieren die Schüler zunächst eine 2-jährige schulische Ausbildungsphase mit den Abschlüssen technische bzw. kaufmännische Assistenz sowie der Fachhochschulreife. Integrierte Praktika ermöglichen den Erwerb der erforderlichen Praxisanteile. Bei erfolgreichem Bestehen folgt eine 1,5-jährige betrieblich ausgerichtete Ausbildungsphase als Fachinformatiker/in Systemintegration bzw. als Kaufmann/-frau für Bürokommunikation ohne Berufsschulunterricht mit dem jeweiligen Kammerabschluss. Der Schulversuch wurde an zwei Hamburger Schulen zunächst in jeweils zwei Durchgängen erprobt.

An diesem Setting werden bereits spezifische Problembereiche deutlich, mit denen es bei der Evaluation neuer Bildungsgänge umzugehen gilt: Zum Einen ist die Anzahl der Schüler in diesem neuen Bildungsgang so gering, dass nicht einmal eine Volluntersuchung valide Aussagen möglich macht. In diesem Zusammenhang kann also nicht von einer externen, sondern lediglich von einer internen Validität gesprochen werden (vgl. Kapitel 4.1). Zum anderen ist ein Ziel der bildungspolitisch Verantwortlichen, die Warteschleifen im Übergangssystem zu verringern, wobei gleichzeitig zu klären ist, ob dieses Ziel mit dem Bildungsgang überhaupt erreicht werden kann (vgl. Kapitel 4.2).

3.2 Evaluationsauftrag

Um die Erfahrungen aus dem Schulversuch weit über die Anlage des Schulversuchs, die Bindung in der Region Hamburg und die Einordnung als Maßnahme zur Vermeidung von Warteschleifen in den Fokus zu nehmen, wurde der Schulversuch seit Februar 2009 von einem Konsortium des Instituts für Berufs- und Wirtschaftspädagogik der Universität Hamburg wissenschaftlich begleitet und evaluiert. Als Auftrag für die wissenschaftliche Begleitung wurde in der Leistungsbeschreibung formuliert: „Die wissenschaftliche Begleitung soll die beiden neu strukturierten Ausbildungsformen evaluieren und ihre Erkenntnisse für die Praxis in den Schulen und den Ausbildungsbetrieben nutzbar machen.“ (HIBB 2008, 2). Obwohl die Evaluation somit ausdrücklich zum Ziel hatte, Erkenntnisse zu liefern, die zur Weiterentwicklung des Schulversuchs führten, sollten Eingriffe in das Schul- und Unterrichtsgeschehen sowie insbesondere in die Lehr-Lernprozessgestaltung auf ein Minimum reduziert werden. Unabhängig davon galt es, Informationen darüber zu generieren,

- in welchem Maße und mit welchem Aufwand der Schulversuch die angestrebten Ziele erreicht,
- welche Effekte und Nebeneffekte erzielt werden und
- unter welchen Voraussetzungen der Schulversuch nachhaltig fortgeführt werden oder auf andere Ausbildungsbereiche transferiert werden kann (vgl. GILLEN/ CIRULIES 2010).

Insofern richtete sich das Interesse der Evaluation nicht allein auf die Implementation eines neuen Bildungsgangs (vgl. Kapitel 4.3), sondern auch auf die Erhebung der damit entstehenden Wirkungen und den Innovationsgehalt (vgl. Kapitel 4.4). Es handelt sich also um eine Kombination aus Wirkungs- und Implementationsevaluation (vgl. KROMREY 2003, 19). Aus diesem Grund war für die Intervention eine Kombination aus formativer und summativer Evaluation vorgesehen, die an bestimmten inhaltlichen Schwerpunkten um einen konstruktiven Zugang erweitert werden musste.

3.3 Interventionsstrategie

Daten für eine summative Evaluation werden vor, während und im Anschluss an den Entwicklungsprozess erhoben, um durch die Ergebnisse ein differenziertes Bild zu verschiedenen Zeitpunkten zu erhalten und unter verschiedenen Fragestellungen miteinander zu vergleichen. Summative Fragestellungen im Evaluationsdesign des Schulversuchs betrafen vor allem Lern- und Prüfungserfolge sowie Aussagen zur Akzeptanz und zur Verknüpfung mit dem Arbeitsmarkt. Als summative Datenbasis wurden während des Schulversuchs u.a. Kompetenzmessungen bei den Schülern durchgeführt, die Arbeitgeber schriftlich zur Akzeptanz des Bildungsgangs befragt und Interviews mit den bildungspolitisch Beteiligten durchgeführt. Zum Vergleich und zur Einordnung der Ergebnisse wurde auf Erhebungen der Fachleistungen in Anlehnung an den bestehenden Hamburger Schulleistungstest ULME (vgl. LEHMANN/ SEEBER 2007) zurückgegriffen.

Eine formative Evaluation erhebt Daten während des Entwicklungsprozesses, um mit den Ergebnissen und der Interpretation dieser Daten steuernd in den fortlaufenden Prozess einzugreifen. Sie gestaltet das zu evaluierende Programm oder Produkt mit (KROMREY 2003, 17). Formative Fragestellungen im Evaluationsdesign des Schulversuchs bezogen sich u.a. auf die curriculare und didaktisch-methodische Entwicklung und Implementation der neuen Ausbildungsform in Schulen und Betrieben sowie auf Maßnahmen zur Prozessanalyse und -gestaltung. Die Datenbasis der formativen Evaluation wurde in Beobachtungen und Interviews erhoben, bewertet und über verschiedene Kanäle in den Implementationsprozess rückgeführt.

Das Evaluationsdesign folgt einem handlungsorientierten Ansatz, der „flexibel, adaptiv und gerade für komplexe, unstrukturierte Situationen sehr geeignet“ scheint (WILL/ WINTELER/ KRAPP 1987, 33). Um die an dem Schulversuch Beteiligten möglichst umfassend in die Evaluations- und Konstruktionsprozesse einzubeziehen, wurde für das Evaluationsteam eine

Organisationsform aus abgeordneten Lehrern und Konsultanten gewählt, die es möglich machte, auch während des Prozesses schnell auf die Beteiligten zuzugehen und sie zu informieren (vgl. Abb. 1).

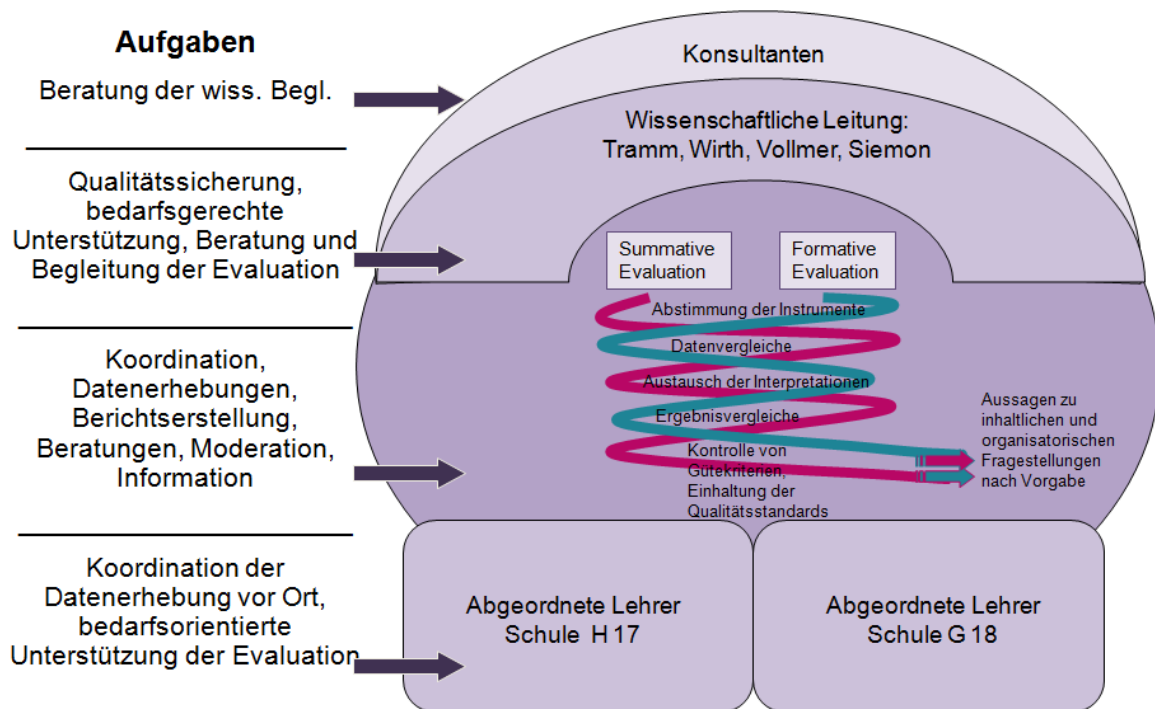


Abb. 1: Organisation der wissenschaftlichen Begleitung

In diesem Sinne wurden für den Schulversuch Lehrer aus den beteiligten Schulen für die wissenschaftliche Begleitung teilabgeordnet, um so eine Beteiligung auch organisatorisch im Evaluationsdesign zu implementieren. Die Aufgaben der abgeordneten Lehrer an dieser evaluatorischen Schnittstelle umfassten dabei sowohl

- allgemeine Aufgaben wie die Pflege und Gestaltung des Informationsflusses zwischen Evaluationsteam und jeweiliger Schule, die Gestaltung der Zusammenarbeit zwischen Schule und wissenschaftlicher Begleitung sowie die beschreibende und rekonstruktive Dokumentationen des Verlaufs des Schulversuches in ausgewählten Handlungsfeldern (Auswahlverfahren, Praktika, schulischer Ausbildungsverlauf, Prüfungsgestaltung, Curriculumentwicklung) als auch
- Mitarbeit bei Maßnahmen, die von der Universität initiiert werden, u.a. Terminkoordinierungen für die Datenerhebung, Information der beteiligten Lehrerteams über Maßnahmen der Evaluation, Begleitung von Veranstaltungen (Fachtagungen, Erfahrungsaustausche, Workshops) sowie konkrete Berichtslegung über Projektverläufe und
- Mitarbeit bei Maßnahmen, die von der Schule initiiert werden. Dazu gehören v.a. Ermittlung und Weiterleitung von Beratungsbedarf, Information des Evaluationsteams über aktuelle Entwicklungen in den Lehrerteams und in den Klassen, Initiierung von

schulinternen bzw. schulrelevanten Veranstaltungen unter Einbeziehung der wissenschaftlichen Begleitung und Mitgestaltung bei der Curriculumentwicklung, der Lernortkooperation und bei der Integration der Fachhochschulreife.

Der wissenschaftlichen Begleitung gehörten neben den Lehrern und der wissenschaftlichen Leitung auch externen Konsultanten an, die bei speziellen Themenstellungen herangezogen wurden. Darüber hinaus wirkten die Evaluatoren darauf hin, dass die am Schulversuch beteiligten Gruppen in konstruktiver Auseinandersetzung die Rahmenbedingungen des Schulversuchs konkretisieren und fortschreiben (vgl. TRAMM et al. 2009). Die organisatorische Anlage und Struktur der Evaluation schöpfte damit die verschiedenen Interventionsstrategien, die für Evaluationen zur Verfügung stehen, in vollem Umfang aus und bettete sie in ein Forschungsdesign ein, das qualitative, quantitative und konstruktive Verfahren umfasst.

3.4 Forschungsdesign

Für das Forschungsdesign bei der Evaluation des Schulversuchs EARA wurde eine Kombination qualitativer und quantitativer Verfahren der Datenerhebung und -auswertung gewählt. Neuere Ansätze unter dem Namen Mixed Methods unterscheiden sequenzielle (sukzessive) und gleichzeitige (komplementäre) Nutzungsweisen dieser Verfahren (vgl. KUSS 2010, 120).

Eine sequenzielle Anwendung qualitativer und quantitativer Methoden liegt z.B. vor, wenn eine Vorstudie qualitativ, die Hauptuntersuchung aber quantitativ angelegt ist. Für die Evaluation des Schulversuchs EARA wurden qualitative und quantitative Verfahren sukzessive eingesetzt, indem zunächst qualitative Daten erhoben werden, um zu Aussagen zu kommen, die anschließend für die Gesamtheit quantitativ nachgeprüft werden, z.B. Antworten zur Situation der betrieblichen Praktika in den qualitativen Gruppeninterviews der Schüler im Mai 2010, die Eingang in den quantitativen Fragebogen zum Verhältnis von Schule und Betrieb erhielten.

Gleichzeitigkeit wird unterstellt, wenn qualitative und quantitative Methoden kombiniert werden, um entweder die in unterschiedlichen Verfahren gewonnenen Daten gegenseitig zu validieren oder durch Betrachtung unterschiedlicher Perspektiven auf den Gegenstand den Untersuchungshorizont zu erweitern. Eine gleichzeitige Nutzung zielt darauf ab, die Validität der gewonnenen Daten z.B. durch Triangulation zu erhöhen. Triangulation kann die Erhebung von Daten aus unterschiedlichen Quellen, zu unterschiedlichen Zeitpunkten, durch verschiedene Personen oder an verschiedenen Orten umfassen. Es handelt sich nach gängiger Meinung allerdings auch um Triangulation, wenn unterschiedliche Methoden, Beobachter oder Theorien kombiniert werden (FLICK 2009, 310ff.). Ein weiterer Vorteil der komplementären Nutzung verschiedener Forschungsstrategien liegt darin, dass der Blick auf den zu evaluierenden Gegenstand aus verschiedenen Perspektiven vorgenommen und damit komplettiert wird. Bei der Evaluation des Schulversuchs EARA wurden die Schüler zu verschiedenen Zeitpunkten und Inhalten sowohl qualitativ als auch quantitativ befragt, um darüber z.B. zu gesicherteren Kenntnissen über die beruflichen Einstellungen, die Akzeptanz und die Motivation zu gelangen.

Darüber hinausgehend waren konstruktive und entwickelnde Aufgaben explizit Teil des Evaluationsauftrags. Dazu wurden u.a. auf Ansätze des Design-Based Research zurück gegriffen, bei denen der Stellenwert des Designs im Forschungsprozess explizit thematisiert wird, um darüber zu Qualitätskriterien für diesen Prozess zu gelangen. Das forschungsstrategische und –methodische Vorgehen bei der Begleitung des Schulversuch EARA stellte einen interventionsorientierten Einsatz sowie ein iteratives Vorgehen in den Vordergrund.

Nachdem in diesem Kapitel die Aufgaben bei Evaluation und Begleitung neuer Bildungsgänge am Beispiel des Hamburger Schulversuchs EARA kurz skizziert wurden, sollen im Folgenden die Herausforderungen im Hinblick auf den Gegenstandsbereich, die Anspruchsgruppen, den Prozessablauf sowie den Innovationsgehalt diskutiert werden, die es für eine Qualitätssicherung zu bewältigen gilt.

4 Herausforderungen bei der Qualitätssicherung

4.1 Gegenstandsbereich

Für die Evaluation des Schulversuchs definierte die wissenschaftliche Begleitung Schwerpunkte des Gegenstandsbereichs auf drei Ebenen (vgl. Tab. 1): Zum Einen gilt es, die Strukturen des Bildungsgangs zu betrachten, die auch die bestehenden bildungspolitischen, schulorganisatorischen und zielgruppenspezifischen Rahmenbedingungen umfassen. Zum Zweiten werden die Prozesse des Bildungsgangs sowie die in den Prozessen auftretenden Besonderheiten des Bildungsgangs in den Blick genommen und in einem dritten Zugang die Effekte des Schulversuchs erhoben.

Tabelle 1: **Strukturen, Prozesse und Effekte im Schulversuch EARA**

Strukturen	<ul style="list-style-type: none"> • Kooperation von Schule und Betrieb • Schulorganisation • Passung der Zielgruppe • Bildungspolitische Absichten und Bedenken
Prozesse	<ul style="list-style-type: none"> • Relationierung schulischer und betrieblicher Ausbildungsanteile • Curricularer Entwicklungsprozess und curriculare Konzeption • Integration der Fachhochschulreife • Gestaltung der Prüfung
Effekte	<ul style="list-style-type: none"> • Kompetenzen der Schüler • Motivation, Interessen, Wahrnehmungen der Schüler • Vermittlung und Verbleib der Schüler • Imageeffekte und Akzeptanz bei den Anspruchsgruppen • Bildungspolitische Bewertung

Für eine erste Bestandsaufnahme wurden die bestehenden Strukturen zur Kooperation der beteiligten Schulen mit den Betrieben, zur Schulorganisation sowie die bildungspolitischen Absichten analysiert. Weiterhin war eine Fragestellung der Auftraggeber, in wie weit der Schulversuch die beabsichtigte Zielgruppe erreicht. Bei der Frage nach den Strukturen han-

delte es sich um einmalige Erhebungen, die den Status Quo ermitteln und festhalten sollten. Dazu sind insbesondere qualitative Erhebungen eingesetzt worden. So wurden z.B. für die bildungspolitischen Absichten der Akteure halbstrukturierte Interviews durchgeführt und mit der Methode der Qualitativen Inhaltsanalyse (vgl. MAYRING 2008) analysiert und ausgewertet.

Diese Strukturen konnten zunächst als einmalige Daten erhoben werden, sie haben allerdings langfristig Prozesscharakter, wie die Einbettung des Schulversuch EARA in die Prozesse der Hamburger Schulreform deutlich macht. Zu Beginn des Schulversuchs waren einige der umgesetzten Maßnahmen, z.B. die Möglichkeit zur Mehrfachqualifizierung, noch Gegenstand der bildungspolitischen Diskussion. Im Reformprozess an Hamburger Schulen wurde allerdings die Möglichkeit, neben einem beruflichen auch einen allgemeinbildenden Fachhochschulabschluss zu erlangen, als Option auch in andere Bildungsgänge verankert (vgl. HIBB 2011). Im Gegensatz dazu besteht für die berufliche Mehrfachqualifizierung mit einem länderspezifischen schulischen Assistenzabschluss und einem dualen Ausbildungsabschluss weiterhin kaum Akzeptanz bei den bildungspolitisch relevanten Anspruchsgruppen.

Hinsichtlich der Prozesse konzentrierte sich die wissenschaftliche Begleitung auf die Evaluation und konstruktive Gestaltung von Besonderheiten des Bildungsgangs wie z.B. die Integration der Fachhochschulreife, die Verknüpfung schulischer und betrieblicher Anteile, die Integration von Ausbildungen nach Landesrecht bzw. BBiG und die Gestaltung der Prüfungen. Ein weiterer Schwerpunkt lag auf der Konstruktion des intendierten und die Rekonstruktion des realisierten curricularen Entwicklungsprozesses. Bei der Betrachtung der Prozesse im Schulversuch galt es als qualitativ bedeutsam, dass die Anlage des Schulversuchs kohärent aufgebaut war. So konnten z.B. die geforderten betrieblichen Praktika nicht schulisch begleitet werden, weil die Zeiten, in denen die Praktika absolviert werden mussten, in den Schulferien lagen. Qualitative und quantitative Erhebungen in diesem Teil der Evaluation hatten, wenn sie notwendig waren, lediglich explorativen Charakter und bezogen sich darauf festzustellen, welche Vorstellungen die Schüler z.B. in Bezug auf ihren Wunschberuf hatten.

Schließlich erfolgte eine Analyse der Effekte in Bezug auf die Kompetenzentwicklung und die Prüfungsnoten der Schüler, auf die Interessen und Motivation der Beteiligten, die Auswirkungen auf den Ausbildungs- und Arbeitsmarkt einschließlich der Akzeptanz bei den Anspruchsgruppen sowie des Übergangs und der Vermittlung der beteiligten Jugendlichen. Dazu wurden qualitative und quantitative Verfahren genutzt.

Für die Messung der Fachkompetenzen wurde auf die bereits bestehenden und in Hamburg eingesetzten Fachleistungstests ULME III zurückgegriffen, in denen die "im schulischen Kontext erworbene berufliche Fähigkeiten" (LEHMANN/ SEEBER/ HUNGER 2007, 15) erhoben werden. ULME III hat den Anspruch, sich im Gegensatz zu den zentralen Kammerprüfungen „an geltenden Standards der Leistungsdiagnostik [zu] orientieren“ (LEHMANN/ SEEBER/ HUNGER 2007, 16). Diese Instrumente wurden in Abstimmung mit der Abteilung Qualitätsentwicklung und Standardsicherung des Landesinstituts für Lehrerbildung und Schulentwicklung (LIQ) an die Erfordernisse des Schulversuchs EARA angepasst.

Gleichwohl ergaben sich nach Einsatz der modifizierten ULME III- Tests weitere Schwierigkeiten, die mit der geringen Stichprobengröße zusammenhingen. Klassische testtheoretische Gütekriterien gehen von einer experimentellen Anlage der Untersuchung aus. Dabei wird aus einer Grundgesamtheit eine Stichprobe gezogen und in einem künstlich angelegten Szenario unter Ausschaltung aller anderen Variablen eine Versuchssituation hergestellt. Die Stichprobe muss in dieser Versuchsanlage die Grundgesamtheit repräsentieren (BORTZ/ DÖRING 2005, 398). Eine Bedingung dafür ist, dass die Stichprobe groß genug ist. Bei der Evaluation neuer Bildungsgänge gelten andere Bedingungen: So wird der Eingriff in der Regel nicht in einer Versuchssituation vorgenommen. Stattdessen wird ein neuer Bildungsgang konstruiert und anschließend in der Praxis eingesetzt. Aussagen über den neuen Bildungsgang können dann nur die unmittelbar Beteiligten (Schüler, Lehrer, Ausbilder etc.) treffen, deren Anzahl meist gering und deren Voraussetzungen und Einstellungen meist heterogen sind. So können die Beteiligten zwar in einer Vollerhebung befragt werden, die Ergebnisse dieser Erhebung sollten allerdings nur in Kombination mit komplementär gewonnenen Daten zu einer Aussage über den Erfolg eines Bildungsgangs herangezogen werden. Im Schulversuch EARA wurden in jedem Jahrgang nur jeweils eine Klasse im technischen und eine Klasse im kaufmännischen Bildungsgang des Schulversuchs aufgenommen. Daraus ergaben sich Schülerzahlen, bei denen valide Aussagen zu den Effekten des Schulversuchs aufgrund der geringen Stichprobengröße nicht möglich waren (vgl. Tab. 2).

Tabelle 2: Schülerzahlen im Schulversuch EARA

		Ausbildungsgang begonnen	Schulischen Teil abgeschlossen	Im betrieblichen Teil der Ausbildung
Durchgang 08	Kaufm.	31	20	13
	Techn.	27	15	8
Durchgang 09	Kaufm.	55	32	12
	Techn.	27	11	4

Aus diesem Grund wurden weitere, für die Effekte des Schulversuchs relevante Daten wie z.B. die Vermittlungsquoten und Verbleibsuntersuchungen als Ergänzung für Aussagen zu den Effekten des Schulversuchs herangezogen, obwohl auch diese allein streng genommen den Gütekriterien Validität, Reliabilität und Objektivität nicht genügen.

Vergleichsdaten aus den Jahrgängen der EARA-Schüler zu erhalten, war auch deshalb mit Schwierigkeiten verbunden, weil der neue Bildungsgang eine berufliche und allgemeinbildende Doppelqualifizierung anstrebt, die bis dahin in Hamburg nur getrennt erworben werden konnten. Die Erlangung der Fachhochschulreife ist in Hamburg über einjährige Fachoberschulen bzw. eine zweijährige Höhere Handelsschule möglich. Diese Qualifizierungswege

stehen Jugendlichen aber nur unter bestimmten Voraussetzungen (z.B. dem Notendurchschnitt der vorangehenden Qualifizierung) als Angebot zur Verfügung. Die Eingangsvoraussetzungen der neuen Bildungsgänge unterscheiden sich, bildungspolitisch durchaus intendiert, von den bisher vorhandenen Qualifizierungsangeboten, so dass ein direkter Vergleich der jeweiligen Absolventen immer mit Einschränkungen verbunden war.

Eine weitere Unterscheidung betraf den Vergleich mit den zugrunde liegenden Assistenzberufen: In Hamburg werden für Schüler kaufmännischer und technischer Bildungsgänge unterschiedliche Qualifizierungsangebote angeboten. So ließen sich die Schüler im mehrfachqualifizierenden Bildungsgang TAISI zwar mit denjenigen des Assistenzberufs TAI vergleichen. Für den kaufmännischen Bildungsgang wurde und wird allerdings in Hamburg keine vergleichbare zweijährige vollzeitschulische Assistenzausbildung angeboten, so dass ein direkter Vergleich entfallen musste.

Weiterhin hatten fast alle Bildungsgänge, die zum Vergleich herangezogen werden konnten, unterschiedliche Ausbildungsdauern (vgl. für die folgenden Ausführungen Abb. 2), so dass ein Vergleich entweder nur zeitlich versetzt oder mit unterschiedlichen Eingangsvoraussetzungen verbunden war.

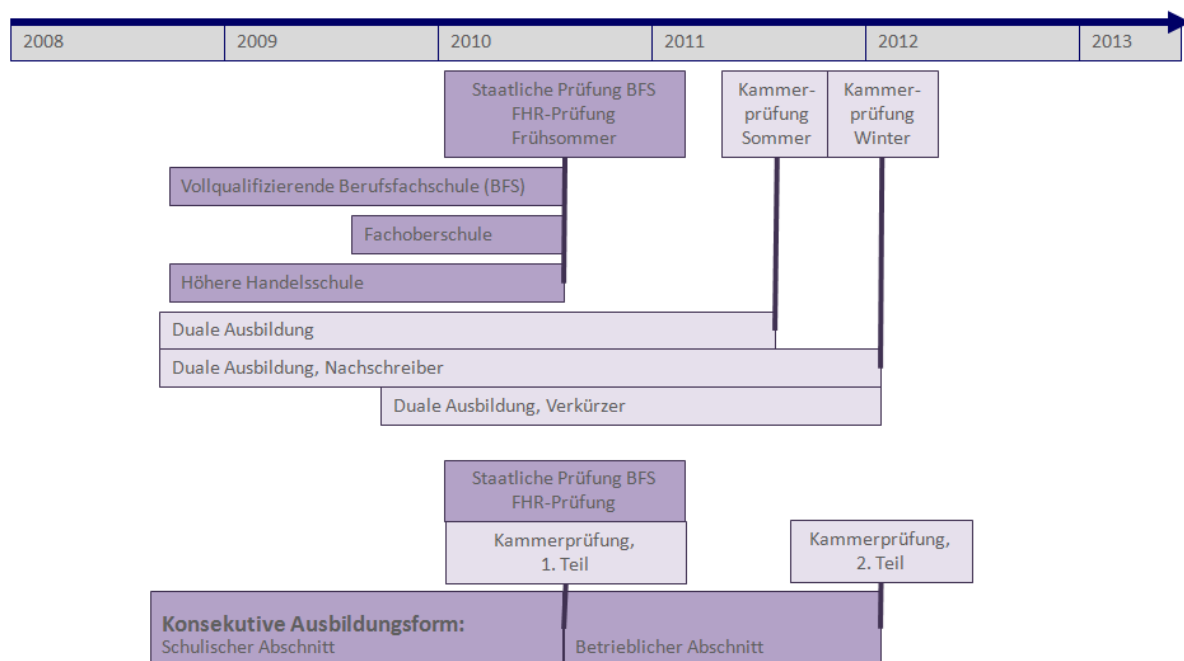


Abb. 2: Ausbildungsformen im zeitlichen Vergleich

So konnten die Schüler der EARA-Bildungsgänge z.B. nicht ohne Weiteres mit den Schülern der jeweiligen dualen Ausbildungsgänge verglichen werden. Diese Vergleichsgruppen hatten entweder ihre Abschlussprüfungen bereits ein halbes Jahr vor den EARA-Schülern absolviert, so dass verschiedene Erhebungszeitpunkte notwendig gewesen wären. Bei denjenigen Absolventen, die zeitgleich mit den EARA-Schülern die Kammerprüfungen absolvierten, handelte es sich entweder um besonders gute Schüler, die ihre

Abschlussprüfung vorgezogen hatten oder es handelt sich um Schüler, die ihre Abschlussprüfung wegen Nichtbestehen eines oder mehrerer Teile wiederholen mussten. Ein Vergleich mit den an den beteiligten Schulen ausgebildeten dualen Ausbildungsklassen, den die Schulen durchaus gewünscht hätten, war aus diesem Grund ebenfalls nicht möglich.

Als Vergleichsgrößen für die berufliche Qualifizierung boten sich andererseits die im Jahr 2005 von ULME III erhobenen Daten an. In ULME III wurden die Daten aller Schüler in den Abschlussklassen von siebzehn Ausbildungsberufen erhoben (Lehmann/ Seeber/ Hunger 2007). So kann zwar davon ausgegangen werden, dass die Vergleichswerte, z.B. der durchschnittlich erreichte Wert im Fachleistungstest, für die Gruppe der erhobenen Schüler durchaus repräsentabel ist, allerdings liegen die ULME III-Erhebungen im Vergleich mit dem Schulversuch EARA auch bereits sechs Jahre zurück.

Als Schwierigkeit bei der Evaluation neuer Bildungsgänge kommt außerdem hinzu, dass von den Ergebnissen, die einzelne Personen im Verlauf des Bildungsgangs erzielen, nicht ohne weitere Daten auf den Erfolg des Bildungsgangs geschlossen werden kann, da die Ergebnisse bei Prüfungen und Tests zunächst Effekte der Lehr-Lernprozessgestaltung sowie der Kommunikation und Interaktion im Unterricht darstellen, statt mit dem Bildungsgang, d.h. mit der (schul-)organisatorischen Ausgestaltung und der Einbettung in das Ausbildungssystem, in Verbindung zu stehen.

Zusammengefasst kann konstatiert werden, dass im Zusammenhang mit dem Gegenstandsreich Fragen zur Aussagekraft der Daten und zur Validität im Vordergrund standen.

4.2 Anspruchsgruppen

SONNTAG (1987, 62) nennt als „Akteure“ für Modellversuche die drei Gruppen Administration, Praxis und Wissenschaft. Für eine umfassende Berücksichtigung der an der Evaluation und Implementation neuer Bildungsgänge beteiligten Personen und Gruppen ist diese Dreiteilung zu undifferenziert. So kann die Anspruchsgruppe Praxis im dual ausgelegten Berufsbildungssystem mindestens in Beteiligte aus Schulen und Betrieben untergliedert werden. Die Akteursgruppen sind weiterhin nach ihren pädagogischen und bildungspolitischen Intentionen zu unterscheiden (vgl. SONNTAG 1987, 64). So finden sich bei den beteiligten Interessenvertretungen mit eindeutig bildungspolitischen Vorstellungen möglicherweise eine andere Schwerpunktsetzung als bei den beteiligten Ausbildern und Lehrern, die ein eher pädagogisch ausgerichtetes Problembewusstsein vertreten. Im Schulversuch EARA wurden die Schüler und gegebenenfalls die Eltern als weitere Anspruchsgruppe in die Erhebung einbezogen. Für die Evaluation neuer Bildungsgänge gilt es daher, mindestens zwischen Zielgruppe (Schüler bzw. Eltern), Umsetzenden (Lehrer, Ausbilder), umsetzender Instanz (Schulen, Betriebe), Interessensvertretungen (in diesem Fall der Landesausschuss für Berufliche Bildung Hamburg) und weiterer Öffentlichkeit (Senat, Presse, wissenschaftliche Öffentlichkeit etc.) zu unterscheiden und vom konkreten Auftraggeber für die Evaluation abzugrenzen.

Eine differenzierte Unterscheidung der Akteure wurde unter anderem notwendig, weil sich im Hamburger Schulversuch immer wieder die Frage danach stellt, ob durch das Angebot die intendierte Zielgruppe erreicht wird. Intendiert war, den neuen Bildungsgang als Angebot für so genannte Marktbenachteiligte auszurichten (vgl. WIRTH 2011). Als marktbenachteiligt gelten Jugendliche, wenn sie „trotz vorhandener Ausbildungsreife und mehrfacher Bewerbungsversuche keinen betrieblichen Ausbildungsplatz auf dem ersten (regulären) Ausbildungsmarkt gefunden haben“ (BSB 2010, 13). Eine vornehmliche Aufgabe der Evaluation dieses neuen Bildungsgangs bestand daher darin, zu untersuchen und zu bewerten, ob der zu erprobende Bildungsgang seinen zum Teil auch kontroversen Zielen und Ansprüchen gerecht wurde, die unterschiedliche Akteure und Anspruchsgruppen an ihn stellen. Für die Frage, ob der Bildungsgang von der erwünschten Zielgruppe gewählt wird, wurden unterschiedlich erhobene Daten ausgewertet. So wurden z.B. Bewerbungsakten (Bewerbungsschreiben, Mitschriften von Bewerbungsgesprächen) einer Inhaltsanalyse unterzogen und die Schüler zu einem Zeitpunkt, an dem sie die Ausbildung bereits begonnen hatten, u.a. nach ihrem Bewerbungsverhalten, ihren Wunschberufen und nach einem von der Handelskammer entworfenen Test (INTAS-Test) befragt. Es stellte sich heraus, dass die Jugendlichen den Bildungsgang unter anderem gewählt hatten, weil sie sich von der Mehrfachqualifizierung Vorteile für ihre Stellung auf dem Arbeitsmarkt versprochen und das Kriterium der Marktbenachteiligung nur eines unter mehreren war. Mit diesem Ergebnis, das weder aufgrund der Größe der untersuchten Zielgruppe noch aufgrund der mangelnden Vorhersagekraft als valide bezeichnet werden kann, ist ein weiteres Qualitätskriterium bei der Evaluation neuer Bildungsgänge angesprochen. Es geht um die Frage, in wie weit gemessen werden kann, ob die Implementation neuer Bildungsgänge zur Lösung vorhandener Probleme beiträgt.

Um den verschiedenen Anspruchsgruppen gerecht zu werden, sind von der Evaluation also Fragen der Zielerreichung und der Vollständigkeit zu thematisieren.

4.3 Prozessablauf

Eine Evaluation betrifft im Idealfall den gesamten Prozess des zu evaluierenden Programms oder Projekts. Nach STOCKMANN (2007, 32) umfasst dieser Prozess die drei Hauptphasen Planung, Implementation und Nachhaltigkeit. Im Projekt EARA wurde nur die Implementationsphase evaluiert und begleitet. Für die Evaluation und Begleitung dieser Phase wurde ein iteratives evaluatives Stufenmodell entwickelt, bei dem auf Anforderungen innerhalb des Implementationszeitraums flexibel reagiert werden konnte (vgl. für die folgenden Ausführungen Abb. 3).

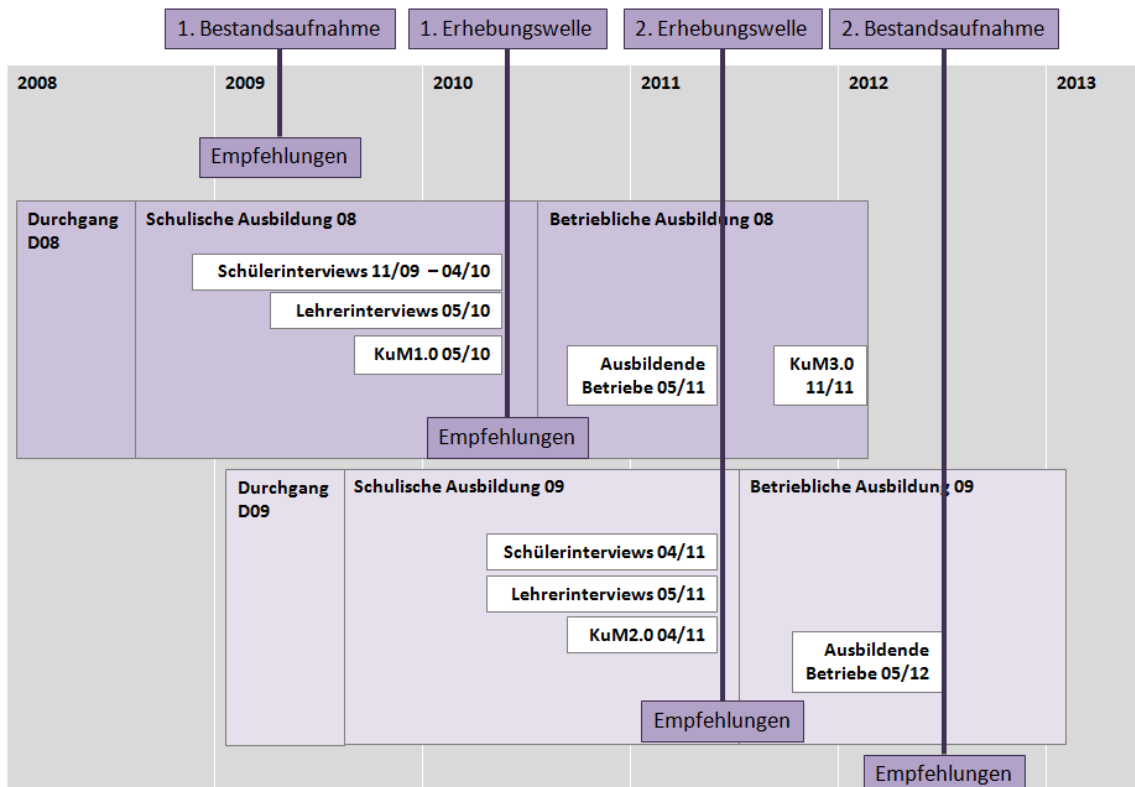


Abb. 3: Iterative Stufenevaluation

Der Schulversuch war in den zwei Durchläufen D08 und D09 vorgesehen, von denen der erste zu Beginn der Evaluation bereits geplant und begonnen worden war. Zu Beginn der Evaluation ab Frühjahr 2009 wurden daher der Stand der curricularen und didaktisch-methodischen Entwicklungsarbeit, Daten zum Auswahlverfahren der Schüler und Schülerinnen sowie die Intentionen, Ziele und Pläne der am Schulversuch Beteiligten in einer ersten Bestandsaufnahme nachträglich erhoben und ausführlich dokumentiert. Aus dieser Bestandsaufnahme der Strukturen resultierten erste Empfehlungen (vgl. Abb. 3), z.B. zum Auswahlverfahren der Schüler an den Schulen, die bereits für den Durchgang 09 berücksichtigt werden konnten.

Eine erste Erhebungswelle im Frühjahr 2010 (vgl. Abb. 3) brachte die qualitativ und quantitativ angelegten Daten für die Beurteilung der Prozesse sowie die Messung erster Effekte des Schulversuchs mit dem Test KuM1.0, in den Items aus ULME III zur Erhebung von Fachleistungen, Motivation sowie zur Selbst- und Methodenkompetenz aufgenommen wurden. Aufgrund der auf dieser Erhebungswelle beruhenden Empfehlungen der wissenschaftlichen Begleitung wurden z.B. wesentliche Teile der Prüfungsgestaltung verändert. Ergebnisse bei der Befragung der Schüler zur Begleitung der betrieblichen Abschnitte lieferten die Grundlage dafür, den Schulversuch sowohl auf curricularer Ebene zu verändern als auch die Evaluation selbst anzupassen. Curriculare Änderungen im Sinne einer formativ angelegten und eingreifenden Evaluation betrafen die schulische Begleitung des betrieblichen Ausbildungsabschnitts für den Durchgang 09, die im Konzept u.a. auf Reflexion der betrieblichen Erfah-

rungen und dem Einsatz von Blended-Learning Maßnahmen beruhen. Für den Durchgang 08 bereiteten die Schulen während des betrieblichen Abschnitts auf die Kammerprüfung im November 2011 vor. Weiterhin wurde der Fragebogen, der von den Schülern Aussagen zur Begleitung des betrieblichen Abschnitts erheben sollte, im Sinne einer offenen Evaluation (vgl. KROMREY 2003, 18) wesentlich erweitert und differenziert. Eine zweite Erhebungswelle hat im Frühjahr 2011 u.a. diese Veränderungen evaluiert und resultierte in weiteren Empfehlungen an die Berufsbildungspraxis. So wurde z.B. darauf hingewiesen, dass in den Schulen noch immer keine angemessene Vor- und Nachbereitungen der betrieblichen Praktika stattfinden. Im Frühjahr 2012 wird der erste Durchgang des Schulversuchs nach der Kammerprüfung auf den Arbeitsmarkt gelangen. Zu diesem Zeitpunkt ist eine zweite Bestandsaufnahme geplant (vgl. Abb. 3).

Dieses Vorgehen entspricht in seiner Grundstruktur den Vorstellungen des Design-Based Research, bei dem ein Prototyp unter Einbezug der Praktiker entwickelt, in der Praxis eingesetzt und anschließend evaluiert und iterativ verändert werden kann (vgl. APREA 2009). Herausforderungen bei diesem Herangehen liegen darin, die Implementation, Konstruktion und Evaluation neuer Bildungsgänge als einen iterativen Prozess zu behandeln. Fragestellungen zur Qualitätssicherung beziehen sich dann darauf, den Grad der Sättigung festzustellen, d.h. denjenigen Zustand des neuen Bildungsgangs, in dem eine erneute Änderung nicht mehr zu einem wahrnehmbar verbesserten Ergebnis führen würde. Alternativ könnten Kriterien dafür formuliert werden, ab wann ein konstruktives Eingreifen in den Prozessablauf nicht mehr zu einem dem Aufwand angemessenen Nutzen führen würde.

Eine weitere Herausforderung in Bezug auf die unterschiedlichen Anspruchsgruppen ist im direkten Zusammenhang mit dem Evaluationsdesign anzusprechen: Im Sinne einer Handlungsforschung oder auch mit Rückgriff auf Ansätze des Design-Based Research wird bei Konstruktion und Implementation eines neuen Bildungsgangs Wert darauf gelegt, alle Beteiligte in den Prozess einzubeziehen. Dazu ist im Verlauf der Evaluation des Schulversuchs EARA unter anderem ein gestuftes Verfahren der Rückmeldung entstanden, das den verschiedenen Anspruchsgruppen in Abhängigkeit von der rückzumeldenden Information benachrichtigt. So werden die Informationen über konkrete unterrichtliche Prozesse direkt an die bei der Lehr-Lernprozessgestaltung Beteiligten (Lehrer und Schüler) rückgemeldet, während Aussagen über Effekte des Schulversuchs zunächst den beteiligten Schulen und anschließend dem Auftraggeber vorgestellt wurden.

Im Zusammenhang mit dem Implementationsprozess beziehen sich offene Fragen zur Qualität also insbesondere darauf, den Sättigungsgrad zu bestimmen und eine iterative Form der Umsetzung für die jeweiligen Evaluationsschritte zu finden.

4.4 Innovationsgehalt

Der Schulversuch EARA setzte sich zum Ziel, Jugendlichen den Übergang von allgemeinbildenden Schulen in die Ausbildung zu erleichtern. Mit diesem Anspruch stand der Schulversuch in einer regionalen und überregionalen bildungspolitischen Diskussion, deren gesetzli-

che Grundlage in der Änderung des Berufsbildungsgesetzes (BBiG) zum 1. April 2005 zu sehen ist. Im neuen BBiG wurde den Ländern in § 43 (2) ermöglicht, Absolventen von Vollzeitschulen künftig per Rechtsverordnung zur Kammerprüfung zuzulassen. Jugendliche, die eine vollzeitschulische Berufsfachschule absolviert haben, können somit zusätzlich ein Zeugnis der Kammern erhalten. Eine weitere Änderung betrifft den § 7, durch den die Länder Verordnungen erlassen können, inwieweit vollzeitschulische Ausbildungen auf eine duale Ausbildung anzurechnen sind. Jugendliche können so nach dem Besuch einer beruflichen Vollzeitschule eine verkürzte duale betriebliche Ausbildung antreten. Mit diesen beiden Änderungen besteht seit 2005 die grundsätzliche Möglichkeit, länderspezifisch so genannte Kombinationsmodelle einzuführen, in denen eine vollzeitschulische Ausbildung mit einer dualen kombiniert wird. Bildungspolitisch werden Kombinationsmodelle kontrovers diskutiert. So gelten sie zwar einerseits als interessante Möglichkeit bei der Suche nach Lösungen, Ausbildungsstrukturen neu zu gestalten, andererseits werden sie aber von den politischen Akteuren (v.a. Ländervertretern und Kammern) nicht akzeptiert (BELLAIRE/ BRANDES 2008).

Als regionaler Bezug des neuen Bildungsgangs war in den Blick zu nehmen, dass das berufliche Bildungssystem in Hamburg derzeit einer strukturellen, inhaltlichen und didaktischen Reform unterzogen wird, als deren Ziel u. a. eine höhere Durchlässigkeit angestrebt wird, um einen verbesserten Übergang von der allgemeinbildenden Schule in den Beruf zu ermöglichen (BSB 2009). Mit dieser Einbindung in den regionalen Schulreformprozess war verbunden, dass sich die Rahmenbedingungen für den Schulversuch EARA änderten. Gleichzeitig ließ sich beobachten, dass Neuerungen, die im Schulversuch EARA erstmals in der öffentlichen regionalen bildungspolitischen Diskussion aufgenommen wurden, über die Grenzen des Schulversuchs hinweg Gültigkeit erlangten. So werden mittlerweile in Hamburg Möglichkeiten zur optionalen Erlangung der Fachhochschulreife auch in anderen Ausbildungsgängen angeboten. Es handelte sich dabei um einen Vorgang, der durch „inkremental-evolutionäre Neuerungen“ (REINMANN 2005, 54) gekennzeichnet ist. In diesem Verständnis vollziehen sich Innovationen nicht unmittelbar und radikal, sondern in kleinen Schritten. Der Schulversuch steht damit in einer regionalen als auch in einer überregionalen bildungspolitischen Diskussion, anhand derer die Ziele, Effekte und der Innovationsgehalt des neuen Bildungsgangs bewertet werden mussten.

Herausforderungen für die Evaluation sind zum Einen damit verbunden, dass die externen Wirkungen der Innovation auf andere Bildungsgänge nicht systematisch und nachweisbar erhoben werden können. Zum anderen haben diese innovativen Maßnahmen langfristigen Charakter, so dass die Auswirkungen der im Schulversuch angelegten Neuerungen auf die zukünftige Struktur des Bildungswesens in Hamburg nicht abgeschätzt werden können.

In diesem Zusammenhang sind es Qualitätskriterien für eine Diffusion (Verbreitung) und für das Einschätzen künftiger Potenziale, die an Maßnahmen im Bildungsbereich angelegt werden müssten, um den Innovationsgehalt der neuen Bildungsgänge und die Vorhersagekraft der bewertenden Evaluation abschätzen zu können.

5 Fazit: Übertrag der Erfahrungen auf die Evaluation neuer Bildungsgänge

In einem Fazit soll nun in fünf Thesen die Übertragbarkeit der bei EARA entwickelten und erprobten Evaluationsformate, -abläufe und -instrumente in den Blick genommen werden, um anzuregen, in wie weit sie auf die Evaluation neuer Bildungsgänge verallgemeinert werden können. Dazu wird insbesondere auf die in Kapitel 4 herausgearbeiteten Qualitätsmaßstäbe Bezug genommen.

- 1) In der Diskussion um die Qualität bei der Evaluation neuer Bildungsgänge stehen insbesondere Maßnahmen der formativen Evaluation in der Diskussion: Positive Erfahrungen wurden im Schulversuch EARA für die Ebene der Unterrichtsprozesse mit der Einrichtung von Arbeitsgruppen zur Curriculumentwicklung (bestehend aus wissenschaftlicher Begleitung und den verantwortlichen Lehrern) gemacht. Auf schulorganisatorischer und bildungspolitischer Ebene konnten positive Effekte mit regionalen Fachtagungen erzielt werden. Vor allem im Hinblick auf das Verhältnis von erforderlichem Aufwand zum erwarteten Nutzen wurden mit der Teilnahme am Schulalltag (z.B. durch Beobachtungen in den Klassen oder bei Lehrerteamsitzungen) keine Qualitätsverbesserungen erzielt.
- 2) Mit dem Organisationsformat, Lehrer in das Team der wissenschaftlichen Begleitung abzuordnen, wurden bei der Evaluation des Schulversuchs EARA ambivalente Erfahrungen gemacht. Die unterschiedlichen Perspektiven von Evaluatoren und zu Evaluierenden spiegelten sich in der Doppelrolle, die die Lehrer explizit übernahmen. Dies führte einerseits zu Rollenkonflikten, gewährleistete letztendlich aber auch die Umsetzbarkeit der iterativ angelegten Evaluation durch kurze, informelle Wege und implizites Expertenwissen aus der Schule „vor Ort“ im Evaluationsteam.
- 3) Für die formative Evaluation neuer Bildungsgänge wurden mit Blick auf die Implementation in die gängige Unterrichtspraxis eine Kombination aus qualitativen und quantitativen Methoden sowie ein iteratives Evaluationsverfahren gewählt. Das Vorgehen bei iterativen Verfahren kann aus Design-Based Research Ansätzen abgeleitet werden. In diesem Zusammenhang müssen Qualitätsmaßstäbe insbesondere für das Definieren kritischer Abbruch- oder Eingriffsfaktoren (Sättigungsgrad) festgelegt werden. Da die Voraussage des Evaluationsverlaufs in iterativen Verfahren nicht immer exakt möglich ist, sind vorab Maßstäbe dafür zu finden, wann bzw. wie lange sich ein Eingriff in der Praxis bewähren muss und welche Kriterien für eine Bewährung gelten.
- 4) Der Auswahl geeigneter Indikatoren für den Erfolg eines neuen Bildungsgangs (z.B. Verbleib, Schulnoten, externe Fachleistungstests) sowie der Planung der Vergleichsgrößen ist besondere Aufmerksamkeit zu widmen, um die Aussagekraft der erhobenen Daten zu gewährleisten. Der Grad der Zielerreichung, nach dem die Qualität innovativer Bildungsgänge schließlich bemessen wird, richtet sich neben dem Verbleib der Schüler auch danach, in wie weit die Bildungsgänge dauerhaft in das Schulangebot integriert werden. Insofern ist neben der Implementation auch immer die Verstetigungsphase bei

der Evaluation zu berücksichtigen, obwohl der Innovationsgehalt sowie die Auswirkungen auf andere Bildungsgänge nicht abschließend erhoben werden können.

- 5) Für Begleitung und Beratung bei der Implementation haben sich für verschiedene Prozessebenen (Unterrichtsprozesse, schulorganisatorische Prozesse, regionale und überregionale bildungspolitische Prozesse) verschiedene Veranstaltungsformen und Rückmelungsprozesse als sinnvoll erwiesen. Je nachdem, auf welcher Ebene Konstruktion, Implementation und Evaluation stattfinden sollen, herrscht ein anderes Interventions- und Innovationsverständnis, das es angemessen zu kommunizieren und zu berücksichtigen gilt. Dabei können je nach Zielstellung, Personenkreis (Adressaten) und Wirkungsgrad kurz-, mittel- und langfristige Rückkopplungsprozesse unterschieden werden. Bei der Evaluation des Schulversuchs EARA hat sich ein stufenförmiges Modell der Rückkopplung bewährt.

Zusammenfassend kann festgestellt werden, dass Kriterien zur Sicherung von Qualität bei der Evaluation neuer Bildungsgänge über die bekannten quantitativen Gütekriterien sowie über die von der qualitativen Forschung ergänzend vorgelegten Kernkriterien hinausgehen müssen. Dabei können die Standards, die von der DEUTSCHEN GESELLSCHAFT FÜR EVALUATION herausgegeben wurden, allenfalls als Richtlinie gelten. Für die Arbeit in Evaluationsprojekten, an die häufig ein doppelter Anspruch von Bewertung und Beratung herangetragen wird, müssen diese Standards erheblich konkretisiert werden (vgl. Abb. 4).

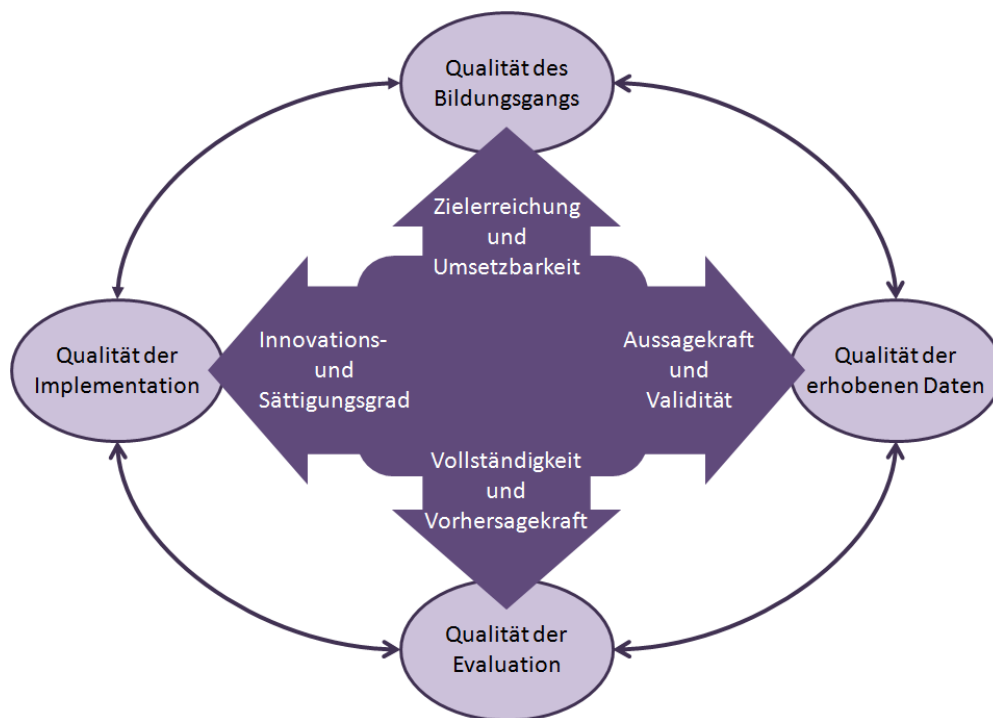


Abb. 4: Qualitätsebenen der Evaluation

Für die Evaluation des Schulversuch EARA wurden in diesem Beitrag Konkretisierungen dieser Standards mit Bezug zu Ansätzen des Design-Based Research vorgestellt und anhand

der dabei aufgetretenen Herausforderungen diskutiert. Für den Gegenstandsbereich standen Fragen der Aussagekraft der erhobenen Daten sowie deren Validität im Vordergrund. Mit Bezug auf die verschiedenen Anspruchsgruppen ließen sich für die Qualität des Bildungsgangs insbesondere Fragen nach dem Grad der Zielerreichung bei der Lösung vorhandener Probleme sowie deren Umsetzbarkeit identifizieren. Beim Gestalten der Prozesse im Wechsel von Implementation, Konstruktion und Evaluation traten Fragen danach auf, nach welchen Abschnitten im Implementationsprozess konstruktiv eingegriffen werden sollte und wann eine Sättigung bei der Verbesserung des Bildungsgangs eintritt. Schließlich bezogen sich Fragen nach der Vollständigkeit und der Vorhersagekraft direkt auf die Evaluation neuer Bildungsgänge.

Literatur

APREA, C. (2009): Design-Based Research. Ein Beispiel aus der universitären Ausbildung angehender Lehrkräfte an wirtschaftsberuflichen Schulen. In: MÜNK, D./ DEISSINGER, T./ TENBERG, R. (Hrsg.): Forschungserträge aus der Berufs- und Wirtschaftspädagogik. Probleme, Perspektiven, Handlungsfelder und Desiderata der beruflichen Bildung in der Bundesrepublik Deutschland, in Europa und im internationalen Raum. Opladen, 107–116.

BELLAIRE, E./ BRANDES, H. (2008): Das duale System flexibel organisieren! Kombinationsmodelle der Ausbildung an Berufsfachschulen und in den Betrieben. In: Berufsbildung in Wissenschaft und Praxis, H. 4, 48-49.

BORTZ, J./ DÖRING, N. (2005): Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler. 3. Aufl. Springer-Lehrbuch. Heidelberg.

BSB, BEHÖRDE FÜR SCHULE UND BERUFSBILDUNG DER FREIEN UND HANSESTADT HAMBURG (2009): Kein Abschluss ohne Anschluss. Übergang von Schülerinnen und Schülern in den Beruf wird grundlegend verbessert. Hamburg.

BSB, BEHÖRDE FÜR SCHULE UND BERUFSBILDUNG DER FREIEN UND HANSESTADT HAMBURG (2010): Ausbildungsreport Hamburg 2010. Hamburg.

DEGEVAL, DEUTSCHE GESELLSCHAFT FÜR EVALUATION (2011): Standards für Evaluation. Online: <http://www.degeval.de/degeval-standards/standards> (11-10-2011).

FISCHER, F./ WAIBEL, M./ WECKER, C. (2005): Nutzenorientierte Grundlagenforschung im Bildungsbereich. In: Zeitschrift für Erziehungswissenschaft, 8, H. 3, 427-442.

FLICK, U.: Triangulation in der qualitativen Forschung. In: FLICK, U./ KARDORFF, E. VON/ STEINKE, I. (Hrsg.): Qualitative Forschung. Ein Handbuch. 7. Aufl. Reinbek bei Hamburg, 309-318.

GILLEN, J./ CIRULIES, N. (2010): Zwischenbericht der wissenschaftlichen Begleitung des Schulversuchs "Erprobung neu strukturierter Ausbildungsformen im Rahmen des Ausbildungskonsenses 2007-2010 (EARA)". Hamburg.

HIBB, HAMBURGER INSTITUT FÜR BERUFLICHE BILDUNG (2008): Leistungsbeschreibung der wissenschaftlichen Begleitung. Hamburg.

HIBB, HAMBURGER INSTITUT FÜR BERUFLICHE BILDUNG (2011): Schulreform. Umfassende Berufsbildungsreform. Online: <http://www.hibb.hamburg.de/index.php/article/detail/194> (11-10-2011).

KROMREY, H. (2003): Evaluierung und Evaluationsforschung: Begriffe, Modelle und Methoden. In: Psychologie in Erziehung und Unterricht, 50, H. 1, 11-26.

KUSS, A. (2010): Mixed Method-Designs : alter Wein in neuen Schläuchen? In: WRONA, T./ FANDEL, G. (Hrsg.): Mixed Methods in der Managementforschung. Wiesbaden, 115-125.

LEHMANN, R. H./ SEEBER, S. (2007): ULME III. Untersuchungen von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Hamburg.

LEHMANN, R. H./ SEEBER, S./ HUNGER, S. (2007): ULME III: Ziele der Untersuchung. In: LEHMANN, R. H./ SEEBER, S. (Hrsg.): ULME III. Untersuchungen von Leistungen, Motivation und Einstellungen der Schülerinnen und Schüler in den Abschlussklassen der Berufsschulen. Hamburg, 15-20.

LÜDERS, C. (2006): Qualitative Evaluationsforschung - Was heißt hier Forschung? In: FLICK, U. (Hrsg.): Qualitative Evaluationsforschung. Konzepte - Methoden - Umsetzung. Reinbek bei Hamburg, 33-62.

MAYRING, P. (2008): Qualitative Inhaltsanalyse. Grundlagen und Techniken. 10. Aufl. Weinheim.

REINMANN, G. (2005): Innovation ohne Forschung? Ein Plädoyer für den Design-Based Research-Ansatz in der Lehr-Lern-Forschung. In: Unterrichtswissenschaft, 33, H. 1, 52-69.

SHAVELSON, R./ PHILLIPS, D./ TOWNE, L./ FEUER, M. (2003): On the Science of Education Design Studies. In: Educational Researcher, 32, H. 1, 25-28.

SONNTAG, K. (1987): Evaluation in der Berufsbildungsforschung im Beziehungsgefüge unterschiedlicher Interessen. In: WILL, H./ WINTELER, A./ KRAPP, A. (Hrsg.): Evaluation in der beruflichen Aus- und Weiterbildung. Konzepte und Strategien. Heidelberg, 61-74.

STEINKE, I. (2006): Gütekriterien qualitativer Forschung. In: FLICK, U./ KARDORFF, E. VON/ STEINKE, I. (Hrsg.): Qualitative Forschung. Ein Handbuch. 7. Aufl. Reinbek bei Hamburg, 319-331.

STOCKMANN, R. (2007): Einführung in die Evaluation. In: STOCKMANN, R. (Hrsg.): Handbuch zur Evaluation. Eine praktische Handlungsanleitung. Münster, 24-70.

TRAMM, T./ VOLLMER, T./ SIEMON, J./ WIRTH, K. (2009): Konzept der wissenschaftlichen Begleitung. Anlage 1 zum Vertrag zur wissenschaftlichen Begleitung des Schulversuchs „Neue Wege zum Ausbildungsberuf“ zw. HIBB und Uni HH/ IBW. Hamburg.

WILL, H./ WINTELER, A./ KRAPP, A. (1987): Von der Erfolgskontrolle zur Evaluation. In: WILL, H./ WINTELER, A./ KRAPP, A. (Hrsg.): Evaluation in der beruflichen Aus- und Weiterbildung. Konzepte und Strategien. Heidelberg, 11-42.

WIRTH, K. (2011): Durchlässigkeit des Bildungssystems durch Dreifachqualifizierung – Chancen für marktbenachteiligte Jugendliche? In: *bwp@* Spezial 5 – Hochschultage Berufliche Bildung 2011, Workshop 02, hrsg. v. RÜTZEL, J./ ZÖLLER, A., 1-12. Online: http://www.bwpat.de/ht2011/ws02/wirth_ws02-ht2011.pdf (11-10-2011).

Dieser Beitrag wurde dem *bwp@*-Format: **DISKUSSIONSBEITRÄGE** zugeordnet.

Zitieren dieses Beitrages

WIRTH, K. (2011): Evaluation neuer Bildungsgänge - Maßnahmen zur Qualitätssicherung zwischen Gegenstandsbereich, Anspruchsgruppen, Prozessverlauf und Innovation. In: *bwp@* Berufs- und Wirtschaftspädagogik – online, Ausgabe 21, 1-21. Online: http://www.bwpat.de/ausgabe21/wirth_bwpat21.pdf (20-12-2011).

Die Autorin



Jun.-Prof. Dr. KARIN WIRTH

Institut für Berufs- und Wirtschaftspädagogik, Universität
Hamburg

Sedanstraße 19, 20146 Hamburg

E-mail: [kwirth \(at\) ibw.uni-hamburg.de](mailto:kwirth@ibw.uni-hamburg.de)

Homepage: www.epb.uni-hamburg.de/node/735