

# Evaluation von Wirkungen – ein Problemaufriss am Beispiel von Wirkkonstellationen interkultureller und interreligiöser Trainings an Schulen

## Evaluation of intercultural and interreligious trainings in schools – a critical depiction of impact research

**Sabine Jungk**

Prof. Dr., Erziehungswissenschaftlerin, Professorin für Interkulturelle Bildung und Erziehung/Sozialpädagogik an der Katholischen Hochschule für Sozialwesen Berlin (KHSB)

**Andrea Nachtigall**

Prof. Dr., Erziehungs- und Politikwissenschaftler\*in, Professor\*in für Theorie und Praxis Sozialer Arbeit mit dem Schwerpunkt Jugend- und Schulsozialarbeit an der Alice Salomon Hochschule Berlin

### **Abstract (Deutsch)**

*Dieser Beitrag schließt an aktuelle Debatten um Wirkungen und Wirkfaktoren in der Jugendbildung an. Empirischer Bezug ist eine von uns 2014 abgeschlossene Evaluation eines interkulturellen und interreligiösen Bildungsprogramms, das seit mehreren Jahren an verschiedenen Berliner Schulen als non-formales Bildungsangebot umgesetzt und weiterentwickelt wird. Davon ausgehend problematisieren wir ein Bildungs- und Evaluationsverständnis, das sich v.a. auf evident mess- und steuerbare Erfolge konzentriert – wie es nicht zuletzt durch Begriffe wie Training und Kompetenz (-steigerung) transportiert wird. Angesichts der Komplexität von Bildungsprozessen, deren „Erfolg“ von vielfältigen Variablen wie z.B. den räumlichen, institutionellen, gesellschaftlichen und persönlich-individuellen Kontextbedingungen abhängt, sind qualitative Verfahren der Beobachtung und Befragungen am ehesten geeignet, um Effekten und Wirkungen pädagogischer Programme mit ihren (nicht-intendierten) „Nebeneffekten“ empirisch auf die Spur zu kommen. Anhand ausgewählter Ergebnisse aus leitfadengestützten Interviews mit Programmteilnehmer\*innen zeigen wir, wie Jugendliche Bildungsinhalte aktiv aneignen und weiterverarbeiten und welche zentrale Rolle dabei die Interaktionsbeziehung zwischen Anleiter\*innen und Teilnehmer\*innen spielt.*

*Schlagwörter: Evaluation, Wirkungen, Jugendbildung, Soziale Arbeit, interkulturelle und interreligiöse Trainings, rekonstruktive Verfahren*

### **Abstract (English)**

*This article examines aspects of impact research, following the findings of an evaluation project concerning intercultural and interreligious trainings in schools. Data shows that the use of qualitative research methods is able to reflect intended and unintended cognitive, emotional and conative learning processes and results in its complexity. Discussing psychometric methods, the authors call into doubt that an evidence based evaluation approach could represent the effects adequately. Furthermore, selected statements of students participating in the programme show the crucial role of interactive effects which call into question the effective teaching model as well.*

*Keywords in English: evaluation, impact, youth education, social work, intercultural and interreligious trainings, qualitative research*

## 1. Einleitung: Evaluation eines interkulturellen und interreligiösen Bildungsprogramms

Evaluationen, insbesondere, wenn sie als Wirkungsforschung angelegt sind, sind heute eingebunden in Macht- und Interessenzusammenhänge, die über die Verteilung knapper gewordener öffentlicher Mittel entscheiden. Dies zeigt sich insbesondere in den Bereichen Bildung und Soziale Arbeit. Vor dem Hintergrund gesellschaftlicher Ökonomisierung, steigendem Effizienz- und Kostendruck, gepaart mit neoliberalen Aktivierungszwängen sind soziale und Bildungs-Träger einem wachsenden Legitimationsdruck ausgesetzt. Vor allem kleinere, projektgebundene Träger sind gezwungen, fortwährend zu „beweisen“, dass ihre Angebote auch wirklich „wirken“ und zum gewünschten „Erfolg“ führen – z.B. durch externe Evaluationen. Dahinter stehen nicht nur kalkulatorische Überlegungen – sondern auch ein gewandeltes Verständnis von funktionaler, output- und leistungsorientierter Pädagogik, Bildung und Sozialer Arbeit, die nun fortwährend in ihren Qualitätszielen bestimmt, in ihrer Erbringung überprüft und ggf. standardisiert in ihren Effekten kontrolliert werden (sollen).<sup>1</sup>

Evidenzbasierten Ansätzen liegt dabei der Glaube an verlässlichen, möglichst eindeutigen Wenn-Dann-Aussagen zugrunde. Was für die Naturwissenschaften noch zutreffen mag, lässt sich auf komplexe soziale und pädagogische Problemzusammenhänge nicht ohne weiteres übertragen: Ein lineares Ursache-Wirkungs-Verständnis verkennt die Komplexität der Kontextbedingungen, vor allem aber den Eigensinn von Lehr-Lern-Prozessen, die von aktiver Aneignung und Umdeutung der Bildungsinhalte durch die Adressat\*innen geprägt sind, und daher per se nur begrenzt kontrollier- und steuerbar sind (zur

Kritik vgl. zuletzt Böttcher/Nüsken 2015, weiterführend z.B. die Beiträge in Otto et al. 2010).

Um Missverständnissen vorzubeugen: Unbestritten ist, *dass* (Jugend-) Bildung wirkt (vgl. für zahlreiche Beispiele Lindner 2015). Zudem ist Nicht-Evaluieren auch keine Lösung, da (externe) Evaluationsinteressen und -anfragen nicht einfach ignoriert werden können, vor allem aber, weil Evaluationen erhebliche Erkenntnisse zutage fördern können, die die Qualität von Sozialer und Bildungsarbeit verbessern helfen.<sup>2</sup> Zu fragen bleibt jedoch, *wie* und *welche* Wirkungen und Wirkfaktoren der Jugendbildung mit welchen empirischen Verfahren erfasst werden können. Für den Bereich der (außerschulischen) Kinder- und Jugendhilfe ist eine systematische, praxisfeldübergreifende Erforschung von Wirkungen bislang nicht erfolgt, wie Lindner (2015:382) konstatiert. Es existiere zudem kein verallgemeinerbares zuverlässiges Forschungsdesign zur empirischen Erfassung der unterschiedlichen Wirkfaktoren (ebd.).

Ausgangspunkt unserer Überlegungen ist ein Evaluationsforschungsprojekt, das wir im Auftrag eines außerschulischen Bildungsträgers 2013/2014 durchgeführt haben. Gegenstand der Evaluation war ein interkulturelles und interreligiöses Bildungsprogramm („Training“), das sich an Schüler\*innen der 7. bis 10. Klasse (Altersspanne 12 bis 17 Jahre) richtete und an verschiedenen Berliner Schulen umgesetzt wurde. Die Workshops fanden als freiwilliges Angebot im Unterricht oder im Rahmen der Schulsozialarbeit, zumeist außerhalb der regulären Unterrichtszeiten (z.B. in Projektwochen) und im Klassenverbund statt; sie boten bewusst Gelegenheiten und Anreize zum non formalen und informellen Lernen. Lehrer\*innen waren von der Teilnahme ausgeschlossen.

Im Folgenden werden wir das Evaluationsprojekt und seine Umsetzung kurz vorstellen (2.). Es folgen methodische Überlegungen (3.), wie sich Wirkungen von pädagogischen Maßnahmen und Bildungsprogrammen überhaupt angemessen erfassen lassen. Dabei werden wir zeigen (4.), dass bereits der Begriff „Training“ irreführend ist und von einer verkürzten sozialtechnologischen Vorstellung von Input-Output-Relationen ausgeht, verbunden mit einer Vermessung des jeweiligen „Erfolgs“. Diese Überlegungen werden wir anhand ausgewählter Ergebnisse der Evaluation illustrieren – und damit den Nutzen qualitativ-rekonstruktiver Methoden unterstreichen (5.). Unser Beitrag schließt mit Anregungen zur kritisch-reflexiven Weiterentwicklung der Wirkungsforschung im Sinne eines demokratischen und empowernden Gebrauchs von Evaluationen (6.).

## **2. Ziele und methodische Umsetzung der Evaluation: qualitative und quantitative Befragung**

Ziel und Zweck des Evaluationsprojekts waren einerseits die Prüfung der Konzeptqualität sowie der Prozesse und Strukturen der praktischen Umsetzung. Andererseits ging es – den Wünschen des Auftraggebers entsprechend – um die Ermittlung von Wirkungseffekten auf Seiten der Zielgruppen, wobei intendierte und nicht-intendierte (positive und negative) Wirkungen erfasst werden sollten.

Zur Beantwortung der Forschungsfragen wurden in einem ersten Schritt die konzeptbezogenen Dokumente theoriegeleitet ausgewertet, um die Angemessenheit des Programms und seiner praktischen, situierten Übungen in Hinblick auf die angestrebten Wissens-, Einstellungs- und Handlungsdimensionen zu überprüfen. Ein weiterer zentraler Schritt zu Beginn der Evaluation war – neben der „Hospitation“ der Evaluatoreninnen bei ausgewählten Trainingssitzungen,

die jedoch nicht systematisch ausgewertet wurden – die Ermittlung zentraler Ziele, Wirkungsannahmen und Erfolgskriterien des Programms zusammen mit dem Projektteam. Ergebnis war eine stärkere Strukturierung und Gewichtung der Ziele durch Identifizierung von Leit- und Feinzielen, die Reduktion auf wenige Hauptziele sowie eine Prüfung der Stimmigkeit zwischen Zielen und Methoden. Die gemeinsam erarbeiteten Erfolgskriterien dienten wiederum als Grundlage für die Evaluation der intendierten Programmziele. Daneben erwiesen sich besonders die Hospitationen als überaus nützlich. Sie ermöglichten nicht nur einen Einstieg ins Feld; die Praxisbeobachtungen lieferten zudem wertvolle Informationen über die Workshopinhalte, den Ablauf, die unterschiedlichen Inputs und interaktiven Übungen sowie v.a. das flexible, situationsangepasste Agieren der Trainer\*innen in der Praxis, das auf Mitwirkung und Einbezug der Teilnehmer\*innen ausgerichtet war. Auf eine systematisierte (teilnehmende) Beobachtung musste jedoch ressourcenbedingt verzichtet werden.

Die empirische Analyse der Effekte und Wirkungen des Programms war als Methodenmix aus qualitativen und quantitativen Elementen in Form von Befragungen angelegt, wobei die qualitative Ausrichtung dominierte. Insgesamt wurden – mit einem zeitlichen Abstand von einigen Wochen – 18 qualitative Interviews durchgeführt, davon sieben Gruppen-Interviews mit je drei Schüler\*innen und elf Einzelinterviews in fünf verschiedenen Klassen an vier Schulen. Befragt wurden insgesamt 28 Schüler\*innen und drei Lehrer\*innen sowie eine Schulsozialarbeiterin. Ergänzend zu den Interviews wurden zwei standardisierte, quantitative Instrumente eingesetzt. Zum einen wurde ein „Ist-Diagnose-Fragebogen“ entwickelt, der jeweils im Vorfeld – im Sinne einer Situationsanalyse – erste

Hinweise auf die jeweilige Relevanz des Themas Religion/Kultur und das damit verbundene Konfliktpotenzial in der Trainingsklasse liefern sollte. Ein zweiter, kurzer „Feedback-Fragebogen“ diente jeweils am Ende eines Workshops dazu, zeitnahe Eindrücke und Reaktionen festzuhalten und lässt Aussagen über die Zufriedenheit und Nutzen allgemein zu.

Insgesamt war die Evaluation partizipativ im Sinne einer engen Zusammenarbeit zwischen Evaluatoreninnen und Projektteam ausgerichtet. So ging es uns v.a. um eine dialogische Reflexion und Weiterentwicklung des bestehenden Angebots, die von grundlegender Wertschätzung getragen und eher im Sinne einer „kollegialen Beratung“ verfährt. Die Instrumente (Fragebogen, Leitfaden) wurden daher ebenso wie Ergebnisse und Zwischenergebnisse an das Team zurückgegeben und gemeinsam diskutiert. Nicht vergessen werden darf dabei jedoch, dass dies für die Projektmitarbeiter\*innen eine zusätzliche Arbeitsanforderung angesichts ohnehin schon knapper zeitlicher und personeller Ressourcen mit sich bringt.

Die Schüler\*innen hingegen wurden aus organisatorischen Gründen nicht direkt in die Gestaltung der Evaluation einbezogen, sondern nur mittels des vorab ausgeteilten „Ist-Diagnose-Fragebogens“ zur interkulturellen und interreligiösen Situation in ihren Klassen befragt. Die qualitativen Interviews (ebenso die einzelnen Workshops) konnten dadurch genauer auf die individuelle Situation in der jeweiligen Klasse und die Schüler\*innen abgestimmt werden.

### **3. Methoden der Wirkungsforschung – eine Problematisierung**

Trainings zu interkultureller Kompetenz sind weit verbreitet, entsprechend werden auch viele Evaluationen in diesem Bereich durchgeführt. Jedoch: Evaluationssettings sind unterentwickelt

und kaum dokumentiert, zumal, wenn es um Wirkungsforschung geht (für studentische Austauschprogramme im internationalen Raum: Stellmanns 2012). Auch wenn man behelfsweise auf verwandte Konzepte bspw. zur Prävention von Vorurteilen, Diskriminierung, Rechtsextremismus ausweicht, wird dieser Mangel kaum behoben. Förderprogramme wie das 2007 gestartete Bundesprogramm „Vielfalt tut gut. Jugend für Toleranz, Vielfalt und Demokratie“ unter der Ägide des Bundesministeriums für Familie, Senioren, Frauen und Jugend sind zwar wissenschaftlich umfangreich begleitet worden (Regiestelle o.J.). Jedoch stehen hier zumeist Fragen der Strukturqualität von Programmen im Mittelpunkt.

Es gilt also, eigene Standards für die Evaluation solcher Programme in Hinblick auf Wirkungen zu entwickeln. Evidenzbasierte Forschung und Evaluation im Bereich sozialer Kompetenzen, und um solche handelt es sich auch bei interkulturellen Kompetenzen, präferieren oft standardisierte, der Testpsychologie entstammende psychometrische Verfahren. Sie gelten als forschungsökonomisch und werden von den einschlägigen Vertreter\*innen häufig auch – neben der direkten Beobachtung – mit der höchsten Reliabilität und Validität ausgestattet.

Als klassischer Ansatz für das Design einer evidenzbasierten psychologisch-pädagogischen Interventionsforschung gelten die Standards: Vortest – Nachtest, Follow up, und Kontrollgruppe (Hager 1998, zit. nach Brohm 2009:215ff.). Dieses Verfahren ist der Logik naturwissenschaftlicher Experimente nachgebildet; jedoch können in (sozial-)pädagogischen Situationen nur quasi-experimentelle Versuchsanordnungen erreicht werden, denn identische Settings sind nicht herstellbar und eine Randomisierung fehlt (ebd.). Weitere evaluationspraktische Schwierigkeiten

entstehen, weil Kontrollgruppen nur schwer zu rekrutieren sind und z.B. die zweimalige Konfrontation von Schüler\*innen mit dem gleichen Erhebungsinstrument zu signifikantem Motivationsschwund führt (v. Saldern 2001).

Unabhängig davon, ob man der wissenschaftstheoretischen Prämisse evidenzbasierter Evaluation folgt oder nicht: Derzeit liegen nach unseren Recherchen keine überzeugenden psychometrischen Instrumente für die benannten Zwecke vor. Selbst in verwandten Bereichen, etwa der Diagnostik sozialer Kompetenzen, stößt man auf das gleiche Dilemma (Kanning 2003:124ff., Brohm 2009:225).

Die Schwierigkeit, zu eindeutig operationalisierten Konstrukten zu kommen, lässt sich beispielhaft an einem der wenigen vorhandenen Instrumente zeigen, dem WIKI-KJ, dem „Würzburger Interkulturelle Kompetenz Inventar für Kinder und Jugendliche“ (Reinders et al. 2011:436). Es umfasst 16 Items zu Verhalten, Einstellungen, Wissen und wird von Schüler\*innen als Selbsteinschätzung beantwortet (vgl. Abb. 1).



Subskala	Items
<b>Kontakthäufigkeit</b>	Kh1 – Wie häufig redest du mit Kindern, die aus einem anderen Land kommen? Kh2 – Wie häufig spielst du mit Kindern, die aus einem anderen Land kommen? Kh3 – Wie häufig besuchst du Kinder zu Hause, die aus einem anderen Land kommen?
<b>Adaptivität</b>	Kf1 – Fühlst du dich sicher, wenn du mit Kindern aus anderen Ländern redest? Kf2 – Ist es für dich einfach, mit Kindern aus anderen Ländern zu reden? Kf3 – Bist du freundlich zu Kindern aus anderen Ländern? Kf4 – Kannst du mit Kindern, die aus einem anderen Land kommen als du Spaß haben?
<b>Offenheit</b>	Ko1 – Ist es für dich okay, mit Schülern aus anderen Ländern für die Schule zu lernen? Ko2 – Willst du mit Schülern aus anderen Ländern gerne etwas zu tun haben? Ko3 – Findest du es gut, gemeinsam mit Kindern aus anderen Ländern nach der Schule etwas zu machen? Ko4 – Findest du es gut, wenn Kinder aus anderen Ländern an deiner Schule sind?
<b>Wissenstransfer</b>	Wt1 – Mein Freund/Meine Freundin erzählt mir, was sie zu Hause gemeinsam machen. Wt2 – Mein Freund/Meine Freundin erzählt mir, wie bei ihm/ihr zu Hause Feste gefeiert werden. Wt3 – Mein Freund/Meine Freundin erzählt mir, was bei ihm/ihr zu Hause gerne gegessen wird. Wt4 – Mein Freund/Meine Freundin erzählt mir, worüber zu Hause geredet wird. Wt5 – Mein Freund/Meine Freundin erzählt mir, welche Regeln in der Familie wichtig sind.

Abb. 1: Würzburger Interkulturelle Kompetenz Inventar für Kinder und Jugendliche (Reinders et al. 2011:436)

Problematisch allein schon ist die Benennung der interkulturellen Dimension mit der Formulierung „Kinder/Schüler aus anderen Ländern“; nicht nur, dass die Schüler\*innen u.U. nicht verstehen, wer hier gemeint ist – es besteht die Gefahr der Reifizierung national gedachter kultureller Unterschiede durch Forschung. Ferner geht die Konstruktion offensichtlich von der Kontakt- bzw. Konfliktthese aus (z.B. Ko3 „nach der Schule etwas gemeinsam machen“). Dahinter steht die Annahme, dass die Wahl von Kontakten nach Herkunft (der Eltern) etwas über interkulturelle Kompetenz aussagt. Dies muss mit Blick auf die in der Forschung relativierte Kulturkonfliktthese zurückgewiesen werden. Kontaktelektivität ist kein „pathologisches Muster“, das „verhindert werden muss“ und nicht per se Ausdruck von Diskriminierung (Dollase 2007:204).

Mit anderen Worten: Das „WIKI-KJ Würzburger Interkulturelle Kompetenz Inventar“ mag reliabel sein, die Validität der Konstruktion sehen wir kritisch. Hinzu kommt, dass Evaluationen mit psychometrischen Instrumenten erfahrungsgemäß eher geringe Effektstärken bei sozialkompetenzbezogenen Interventionen mit präventivem Charakter zeigen (Brohm 2009:213). Wenn Kanning für den Bereich der sozialen Kompetenz konstatiert, dass es ein „allumfassendes Instrument, das sich gleichsam wie der Hase aus dem Zylinder ziehen lässt“ nicht gibt und „sicherlich auch niemals geben“ wird (Kanning 2003:30), da nur schwer Einigkeit über die Konstruktion zu erzielen ist, so gilt dies auch für den Bereich interkultureller Kompetenz.

Einigkeit besteht, dass Wissen, Fähigkeiten und Einstellungen die zentralen Merkmale interkultureller Kompetenz ausmachen – aber das ist eher banal. Eine Selbstbeschränkung auf ein bestimmtes Erfassungsinstrument ist jedoch kaum zu rechtfertigen

(vgl. auch v. Saldern 2001:3). Wir ordnen unser Untersuchungsdesign dagegen der nutzenorientierten Evaluation zu, mit dem Ziel der Erfassung des individuellen Nutzens für die einzelnen Beteiligten (vgl. Soellner 2010). Dies nicht nur, damit die praktische Verwertbarkeit der Evaluationsergebnisse für alle Stakeholder im Mittelpunkt steht, sondern v.a. mit Blick auf die Frage, wie Bildungsprozesse – und um solche handelt es sich gemäß den Absichten interkultureller und interreligiöser Trainings – auf der Wirkungsebene zu untersuchen sind.

#### **4. Training und Bildungseffekte hinterfragen**

Moderne Lerntheorien kritisieren eine simplifizierte Vorstellung von Input-output-Relationen, die einem naiven Verständnis der Evaluation von Wirkungen unterliegen. Sie schließen an das schon von Luhmann konstatierte „Technologiedefizit“ an, nach dem erzieherisches Handeln nicht von außen Menschen von einem Zustand A in einen Zustand B überführen kann (Galuske 2013:64ff.). Sie gehen aber darüber hinaus, indem sie die aktive Rolle von Lehrendem und Lernenden hervorheben: „Lernen ist nicht das abbildhafte Ergebnis von Lehren“, sondern ist „immer eine selbstgesteuerte, soziale und situative Bedeutungskonstruktion des Schülers auf der Basis seiner bisher durch Erfahrung erworbenen Strukturen des Denkens, Fühlens, Wollens und Könnens, wenn es erfolgreich und dauerhaft sein soll“ (Wiater 2010:272). Lerninhalte werden also aktiv angeeignet und umgearbeitet. Ein bestimmtes Bildungsziel, eine erwünschte Verhaltensänderung sind nur zu erreichen, „wenn es der Klient will, wenn er – aus welchen Gründen auch immer – einen Sinn darin sieht, sich ‚auf den Weg‘ zu machen“ (Galuske 2013:50). (Sozial-)Pädagogische Fachkräfte sind jedoch nicht in die Bedeutungslosigkeit verwiesen,

sondern ein\*e Partner\*in im Prozess von Ko-Produktion und Dialog. Im pädagogischen und sozialarbeiterischen Bereich zeichnet sich professionelles Handeln entsprechend gerade dadurch aus, dass Entscheidungen flexibel und situationsabhängig getroffen und Wissen oder Methoden fallspezifisch angewendet werden. Stets muss mit der Eigenaktivität der Adressat\*innen gerechnet werden (Haubrich 2009:449), so dass die Wirkung oder gar der Erfolg (sozial-)pädagogischer Maßnahmen niemals eindeutig vorhersehbar oder planbar sind.

Mehr noch: Die Anerkennung des Lernenden allein kann eine demokratische Pädagogik auch *normativ* begründen – und dies führt unweigerlich zu einer Abkehr eines „effective teaching“-Modells, das lediglich zweckrationale Zielerreichung als Ausweis für Wirksamkeit hält (Terhart 2013:134). Bildungseffekte, die über abprüfbares Wissen hinausgehen, entfalten sich langfristig und eigensinnig. Dies stellt generell den Trainingsbegriff in Frage (s.u.) und hat Folgen für die Evaluation. Brohm reflektiert zudem, dass Kompetenzentwicklung „von außen“ nicht möglich ist: „Die pädagogischen Interventionen haben keinen Zugriff auf die intrapersonalen Transformationsprozesse“ (Brohm 2009:213, ähnlich Kanning 2003:13). Auch Wahl (2002:239) verweist darauf, dass „im Interaktionsbereich lange Lernwege unumgänglich sind“. Trainings könnten „nur ein Anfang sein“ (ebd.), aber umgekehrt sei deutlich: „Ohne jede Form praktischer Übungen bleiben träges Wissen und träge Emotionen was sie sind: Hindernisse auf den Wegen vom Handeln zum Wissen und vom Wissen zum Handeln“ (ebd.:240). Um gezielt „handlungssteuernde Strukturen und Prozesse bewusst und damit bearbeitbar zu machen“ (ebd.), käme es v.a. auf ein wohlbegründetes, konsistentes Übungssetting an. (Dies zu überprüfen, war Gegenstand der theoriegeleiteten Konzeptauswertung, vgl. 2.)

Bildungsprozesse sind einerseits der Introspektion zugänglich (Bildungsromane sind ein Zeugnis dafür). Sie lassen sich andererseits von außen beobachten. An dieser Stelle möchten wir auf einen Vorschlag von Klafki (1994:280ff.) aufmerksam machen. Wenn es um wertorientierte Bildungsprozesse geht, deren Ziele nicht lediglich auf Wissen, sondern auf Handeln, Einstellungen, Haltungen, Überzeugungen gerichtet sind, so Klafki, solle als Erfolgskriterium „Erweisbarkeit“ statt Überprüfbarkeit gelten. Auf Basis von mit den Schüler\*innen zuvor ausgehandelten Kriterien könnten so Effekte durch Beobachtungen ermittelt werden. Dies entspräche einem nicht-standardisierten, vielmehr partizipativ entwickelten und in der Durchführung auf Selbst- und Fremdbeobachtungen setzenden Setting. Ähnlich plädiert auch Kanning (2003) dafür, Kompetenz als stabiles Verhalten, und nicht etwa nur als Disposition, aus „der Beobachtung des Verhaltens über mehrere Situationen hinweg“ zu ermitteln (Kanning 2003:13). Solche ressourcenaufwendigen Langzeitbeobachtungen kommen für externe Evaluationen allerdings kaum in Betracht.

Unsere Erfahrungen zeigen aber, dass qualitative Gruppen- und Einzelinterviews im Sinne eines altersangemessenen, diskursiven und partizipativen Settings sich zur Evaluation sehr gut eignen. Schüler\*innen können dabei das Bildungsprogramm nochmals aktualisieren, reflektieren und in der (Nach-)Erzählung produktiv Bedeutungen für ihre Person hervorbringen.

So stellen die Interviews nicht nur die empirische Basis der Evaluation dar, sondern tragen unmittelbar selbst dazu bei, die Bildungsprozesse auf Seiten der Jugendlichen zu verstärken und zu intensivieren. Im Rahmen der leitfadengestützten, jedoch offen



angelegten Interviews wurden die Workshops von den Schüler\*innen – mit zeitlichem Abstand von einigen Wochen – erneut erinnert und narrativiert. Besonders deutlich zeigen sich die sekundären Lernprozesse in den Gruppeninterviews, in denen die Schüler\*innen über Sinn und Zweck der Workshops dialogisch reflektierten und ihre subjektiven Erinnerungen und Erfahrungen mit den anderen abglichen. Die Schüler\*innen diskutierten dabei nicht nur Inhalte und Ziele des Workshops, sondern auch den individuellen Nutzen – wobei unterschiedliche Erinnerungen, Einschätzungen und Bewertungen der Schüler\*innen bzgl. der Wirkungen des Programms zu Tage traten und erneuter Anlass diskursiver Aushandlungen waren.

## **5. „Ich bin die Wodka“ – ausgewählte Ergebnisse der qualitativen Befragung**

Unsere Evaluation hat gezeigt, dass viele der im Vorfeld gemeinsam festgelegten Fein- bzw. Handlungsziele mit dem Programm erreicht werden konnten. Hier geht es also um die intendierten Wirkungen – die aus der Sicht der Schüler\*innen (mittels qualitativer Befragung) retrospektiv rekonstruiert wurden. Zu den erfolgreich umgesetzten Zielen gehört, dass die Trainings es geschafft haben, neues Wissen zum Thema Kultur und Religion zu vermitteln und für kulturelle und religionsbezogene Zuschreibungen zu sensibilisieren. Die Schüler\*innen stellen in den Interviews mit ausführlichen Schilderungen bestimmter Abläufe während der Workshops dar, wie sie sich z.B. mit Vorurteilen und Stereotypen kritisch auseinandergesetzt und gelernt haben, sich anderen gegenüber respektvoll zu verhalten, ohne Mitschüler\*innen wegen ihrer Religionszugehörigkeit oder kulturellen Herkunft zu diskriminieren. Aus Sicht der Befragten wurden der Zusammenhalt in der Klasse durch die Teilnahme an dem Programm

verbessert und die Konfliktfähigkeit der Schüler\*innen erhöht.

Ein weiterer wichtiger – jedoch nicht explizit intendierter – Effekt der Trainings war, dass die Schüler\*innen sich untereinander besser kennengelernt und neue Dinge voneinander erfahren haben. Offensichtlich haben solche diskursiven, an die eigene Biografie gekoppelten Auseinandersetzungen sonst im Schulalltag, im Unterricht und im informellen Beisammensein, wenig Platz. Auf die offene Einstiegsfrage, was ihnen an den Trainings am besten gefallen habe, antworteten gleich mehrere Schüler\*innen, dass sie im Rahmen des Workshops viel Neues und Persönliches über Ihre Mitschüler\*innen erfahren hätten, wie überrascht sie z.B. waren über die Positionierungen ihrer Klassenkamerad\*innen zu religiösen oder familiären Wertorientierungen. So beschreibt eine Schülerin ihre Erfahrungen folgendermaßen (Hintergrund ist eine Übung mit dem Namen „Identitätskette“): „Mir persönlich hat es so geholfen, zum Beispiel habe ich so erfahren, wie andere aus meiner Klasse denken, was für sie wichtig im Leben ist und so. In dem Sinne war es sehr gut. [...] Ich hatte davor andere Vorstellungen, aber als ich die Ketten von den anderen gesehen habe, da war ich bei manchen überrascht“ (Ab5:220). Diese Übung, bei der die subjektive Wichtigkeit von „Familie“, „Religion“, „Freunden“ etc. durch verschiedenfarbige, auf eine Schnur aufgefädelte Perlen sichtbar zum Ausdruck gebracht wird, ist bei besonders vielen der Befragten positiv Erinnerung geblieben. So ein anderer Schüler: „Erst musst du dich selber einschätzen, dann wurdest du von anderen eingeschätzt und dann hat man halt das den anderen auch vorgestellt und es wurde einfach auch gezeigt, wie sich die anderen fühlen“ (Aa2:187).

Wir haben eine Vielzahl von Belegen aus allen Klassen, dass die Trainings der Kommunikation der Schüler\*innen untereinander neue Impulse geben

und ein intensiver Austausch stimuliert wird, wie er sonst in der Schule kaum zustande kommt: „So hat man mal wirklich über andere Dinge geredet, über die man so gar nicht reden würde. Auf die man auch gar nicht kommt so zu reden, gerade in unserem Alter so irgendwie“ (B3:103). Exemplarisch kann dies an der interreligiösen Zielsetzung verdeutlicht werden. Es ging „wirklich so ins Menschliche rein und ins Religiöse und es war halt sehr tief Sinnig, was in der Schule jetzt nicht so oft da war“ (C3:82), wird aus einer anderen Klasse bestätigt. Die Wirkung sei, „andere Religionen oder Kulturen besser zu akzeptieren, weil ich jetzt auch Hintergründe kenne, warum machen die so etwas. Wenn man dieses Wissen hat, dann weiß man sich auch damit auseinanderzusetzen“ (C2:109). Diese Antworten mögen sozial erwünscht klingen, jedoch bestätigt auch die pädagogische Fachkraft nachdrücklich – und offenbar etwas überrascht – das Interesse der Schüler\*innen: „Da fand ich es ganz spannend, dass manche auch irgendwie so sehr, sehr interessiert waren“ (C1:27) und sich vergleichend mit ihrer familienbiografischen religiösen Prägung auseinandersetzen: „Ah, das ist bei euch auch so“, und dann haben die Schüler auch erzählt: „Ja, bei uns ist es so und so und meine Großeltern haben das noch so und so gemacht“ (C1:37).

Durchgängig konnten wir finden, dass die interviewten Schüler\*innen insgesamt eine sehr wenig problemzentrierte Sicht auf ihr interkulturelles und interreligiöses Miteinander in der Schulklasse haben. Vorurteile und Ausgrenzungen aufgrund von Herkunft oder Religion werden weitgehend negiert, und Stereotype und Vorurteile aufgreifende Redeweisen als „Spaß“ gekennzeichnet: „Zwischen Freunden, da macht man schon ein paar Witze wie ‚Das ist typisch deutsch‘ oder ‚Ja, immer ihr Türken““ (Aa7:21). Eine andere Schülerin russischer Herkunft beschreibt das so: „Ich bin

die ‚Wodka‘, also ich werde inzwischen nicht mehr mit Namen angesprochen von manchen, sondern mit ‚Wodka‘. ((lacht)) Finde ich eigentlich o.k., also ich hab da nichts gegen“ (B4:18). Ironisch oder sarkastisch mit den Provokationen umzugehen, kann wie in diesem und auch dem folgenden Beispiel bis zur spielerischen Selbststigmatisierung führen: „Dann haben wir einen in der Klasse, der ist zu einem Viertel Pole, ne? Und er sagt immer über sich selbst, ich bin ja hier der ‚Klauer““ (B4:22).

Da solche Beleidigungen und stereotypen Redeweisen im Rahmen der Klassengemeinschaft sozial situiert sind, werden sie banalisiert und nicht diskriminierend verstanden. Sie stellen ein Ritual dar; innerhalb dieses „Spiels“ muss die despektierlich behandelte Person gelassen mit Beschimpfungen umgehen und darf keine „Schwäche“ (= Verletzung) erkennen lassen: „Die meisten nehmen solche Beleidigungen mit Humor und kontern halt so, es ist für die ganze Klasse witzig, wenn die es so sehen“ (C3:54).

Allerdings erkennen die Schüler\*innen durchaus, dass manchmal die Grenzen von „Spaß“ überschritten werden – vermeintlicher Spaß kann „auf die Nerven gehen“ oder auch verletzend werden: „Unter Freunden, es sind ja meine engsten Freunde, wenn die das sagen, ist kein Problem. Wenn es dann manchmal zu viel von allen kommt, nervt es natürlich“ (Ab3:22). Ein anderer Schüler formuliert das so: „Ich fand es relativ witzig so, aber die Sache ist halt die, dass es ab und zu mal öfter passiert. Einmal Spaß, zweimal Spaß, beim dritten Mal ist es Spaß, irgendwann wird es zu viel“ (C3:20).

Hier kommt ein (ungeplantes) Ereignis während des Trainings ins Spiel, das einen unintendierten, aber stark prägenden Effekt auslöste. Die Klasse konnte anhand der Reaktion einer der Trainer\*innen stellvertretend miterleben, wie kränkend vermeintliche Witze und Hänseleien sein können.

Eine Trainerin wurde aufgrund ihres Akzentes von einigen in der Klasse verspottet und nachgeäfft; sie zeigte sich verletzt und forderte die Schüler\*innen auf, damit aufzuhören. Diese Situation wird von mehreren Schüler\*innen in verschiedenen Interviews sehr gut erinnert – ohne dass sie von den Interviewerinnen direkt danach befragt wurden – und ausführlich geschildert, was auf einen hohen affektiven Lerneffekt und die wichtige Rolle der ‚Authentizität‘ der Trainer\*innen schließen lässt. So erinnert sich eine Person: „[Name Trainerin] hatte doch einen Akzent. Am Anfang haben sich viele darüber lustig gemacht und sie hat selber sich nachher noch beschwert. Sie meinte ‚Hört auf damit‘. Sie hat selber gezeigt, wie man eigentlich reagieren sollte, wenn so etwas passiert“ (Ab3:202). Auf Nachfrage der Interviewerin: „Und wie war das für euch das zu merken, dass sie das auch nicht so toll fand?“, fährt die Befragte fort: „Ja, peinlich, weil man// Wir haben uns geschämt“ (Ab3:208). Die Klasse habe daraufhin „zwei Tage damit aufgehört“ und sich mit „Kraftausdrücken“ zurückgehalten (Ab3: 272).

Dies sind nur einige Beispiele, die belegen, dass Bildungsprozesse im umfassenden Sinn angestoßen werden. Sie zeigen außerdem, dass auch solche Situationen besonders nachhaltig wirken (oder zumindest in Erinnerung bleiben), die kein fester Programmbaustein waren, sondern die spontan entstanden sind und v.a. mit der Person und Persönlichkeit der Trainer\*innen bzw. der pädagogischen Beziehung zusammenhängen.<sup>3</sup> Eine standardisierte Evaluation könnte solche bedeutsamen Effekte, die auf Kernkompetenzen professionellen (sozial-)pädagogischen Handelns zurückgehen, gar nicht erfassen.

## **6. Fazit: Ein doppeltes Plädoyer für die kritisch-reflexive Weiterentwicklung und den partizipativen und empowernden Gebrauch von Evaluationen**

Vor dem Hintergrund komplexer sozialer Probleme und Ungleichheitsstrukturen einerseits sowie der Vielschichtigkeit von Bildungsprozessen andererseits bleibt ein Wirkungsverständnis, das von monokausalen und eindimensionalen Ursache-Wirkungs-Zusammenhängen ausgeht, unterkomplex und höchst problematisch. Verkannt wird nicht nur die aktive Aneignung und Umdeutung von Lerninhalten, wie sie für Bildungsprozesse zentral sind, sondern auch die prinzipielle Nicht-Steuerbarkeit von Menschen. Mehr noch: (Evidenzbasierte) Ansätze, die v.a. auf wirkungsorientierte Steuerung setzen, scheinen gar von der Hoffnung und dem Wunsch nach der Überwindung des (vermeintlichen) „Technologiedefizits“ geleitet (Albus/ Polutta 2008:262). Auch für die Wissenschaft hat dieses Verständnis Konsequenzen, wie Albus und Polutta kritisch formulieren: „Die Forderung, Forschung darauf auszurichten, zu zeigen ‚was wirkt‘, um Soziale Dienste, Hilfen, pädagogische Prozesse quasi ohne Umwege zu optimieren, stellt eine Abkehr von Forschungsperspektiven dar, die sich z.B. der Rekonstruktion sozialer Praxen widmen, Lebenslagen und -bedingungen sowie Machtverhältnisse analysieren, oder der Angemessenheit sozialer Interventionen nachgehen“ (ebd.: 261). Allein auf messbare Wirkungen und steuerbare (Leistungs-)Erfolge zu setzen, bedeutet damit letztendlich eine problematische Verkürzung und De-Professionalisierung (sozial-)pädagogischer Praxis und Theorie. Stattdessen gilt es Ganzheitlichkeit, Lebenswelt- und Menschenrechtsorientierung sowie kritische (Selbst-)Reflexivität - gegenüber gesellschaftlichen

Machtverhältnissen sowie den eigenen Verstrickungen in diese - als zentrale Professionsmerkmale (neu) zu bestimmen. Professionelles Handeln muss stets am Einzelfall ansetzen, strukturelle Bedingungen sowie individuellen Eigensinn berücksichtigen – was wiederum dem Ruf nach verallgemeinerbaren bildungstheoretischen und methodischen Patentrezepten entschieden zuwiderläuft.

Benötigt werden also andere, angemessenere empirische Konzepte zur Erfassung von Wirkungen und Wirkfaktoren, die jeweils im Einzelfall bzw. unter Berücksichtigung des spezifischen Settings und Kontextes entwickelt werden müssen. Besonders qualitative, rekonstruktive und interpretative Methoden, die nicht nur nach quantifizierbaren Wirkungen, sondern auch den *Wirkungsbedingungen* und *Wirkungsmechanismen*, d.h. den gesellschaftlichen und organisationsbezogenen Rahmenbedingungen und Machtverhältnissen, den ethischen Grundlegungen sowie schließlich den Lebenswelten der Adressat\*innen als entscheidende Kriterien für Wirkungen fragen, sind hier u.E. notwendig (im Anschluss an Albus/ Polutta 2008:263f.). Deutlich zeigt sich in unserer Untersuchung z.B. der Einfluss von „Klasse“ und „Bildungshintergrund“ bei der Rekonstruktion der Wirkdimensionen. Die Fähigkeit, die eigenen Erinnerungen sprachlich wiederzugeben, sowie das Reflexionsniveau bezüglich der Inhalte und Ziele des Programms divergiert stark zwischen den befragten Schüler\*innen des Gymnasiums und der Integrierten Sekundarschule.

Allerdings lassen sich Wirkungen von sozialen oder interkulturellen Trainings, die über eine Wissensaneignung hinaus auch eine Veränderung der Einstellungs- und v.a. Verhaltensebene angestoßen haben, nicht unmittelbar

ablesen oder zweifelsfrei (nur) auf die pädagogische Maßnahme zurückführen. Hier erweist sich die rückwirkende Rekonstruktion von Handlungs- und Wirkungszusammenhängen durch teilnehmende Beobachtung oder Befragungen – möglichst aller Beteiligten – als sinnvoll, um Wirkungen sowie „Gelingensfaktoren“ zu Tage zu fördern. Als ein wichtiges Ergebnis lässt sich z.B. festhalten, dass die Qualität des von uns evaluierten Programms gerade darin liegt, dass die situierten praktischen Übungen in Hinblick auf die angestrebten pädagogischen Effekte theoretisch konsistent sind und die didaktischen Arrangements immer wieder an die vorgefundene Situation angepasst, und damit Selbsttätigkeit und Partizipation der Jugendlichen unterstützt werden.

Um den Nutzen von Evaluation und Wirkungsforschung zu erhöhen, halten wir eine kritisch-reflexive Weiterentwicklung von Evaluationen im Sinne eines demokratischen und empowernden Gebrauchs für notwendig. Besonders ein *partizipatives* Evaluationssetting, das die Mitarbeiter\*innen der zu evaluierenden Einrichtungen und Projekte wie ggf. auch die Nutzer\*innen/Adressat\*innen derselben umfassend beteiligt, in die Gestaltung, Durchführung und Auswertung der Evaluation einbezieht, erweist sich aus mehreren Gründen als sinnvoll. So kann zum einen sichergestellt werden, dass die verschiedenen Perspektiven der unterschiedlichen Stakeholder – wissenschaftliches, praxisbezogenes, lebensweltliches Wissen etc. – nicht gegeneinander ausgespielt, sondern multiperspektivisch verknüpft werden und Synergieeffekte entstehen. Wirkzusammenhänge und Wirkfaktoren können so präziser erfasst werden.

Durch die enge Zusammenarbeit zwischen Evaluatoreninnen und Projektteam lässt sich zum anderen der Kontrollfunktion von Evaluationen

entgegensteuern und der Nutzen für das Projektteam erhöhen. Es zeigte sich, dass bereits die Lern- und Austauschprozesse während der Evaluation ein wertvoller Bestandteil der Evaluation waren; z.B. die gemeinsame Zielermittlung zu Beginn, die als sehr bereichernd wahrgenommen wurde. Evaluationen können in diesem Sinne (für das Projektteam) einen empowernden Nutzen haben.<sup>4</sup>

Roth fordert zudem mit Blick auf Wirkungsforschungen in den „Humandienstleistungsberufen“, dass zumindest diejenigen, „bei denen und mit denen etwas bewirkt werden soll“, angemessen beteiligt werden: „Denn die Zielgruppe, die Betroffenen, Klienten etc. sind es letztlich, die über die Kompetenzen und das Wissen verfügen zu entscheiden, ob eine professionelle Maßnahme etwas gebracht hat. Dazu bedarf es von allen Beteiligten geteilter und transparenter Ziele, die nach Möglichkeit gemeinsam mit der Zielgruppe zu erarbeiten sind. In der Regel sind sie die schwächste Gruppe in professionellen Kontexten. Sie zu stärken, stellt deshalb eine zentrale Bedingung demokratieförderlichen Gebrauchs von Evaluationen dar“ (Roth 2004:9).

Wie die Partizipation und Mitwirkung der Nutzer\*innen bzw. Adressat\*innen von Bildungsprogrammen sichergestellt werden kann, bleibt jedoch vor allem eine Frage der zur Verfügung stehenden Ressourcen. Eine aktive Einbeziehung der Adressat\*innen, hier der Zielgruppe Schüler\*innen, in den gesamten Evaluationsprozess ist zweifelsohne wünschens- und erstrebenswert. Adressat\*innen könnten z.B. bereits im Vorfeld an der Diskussion von Zielen, Wirkungsannahmen und Erfolgskriterien, der Gestaltung der Instrumente (Fragebogen etc.) oder aber an der Diskussion der Ergebnisse systematisch beteiligt werden, wodurch den individuellen jugendlichen Lebenswelten Rechnung getragen würde und Wirkungszusammenhänge

genauer erfasst werden könnten. Besonders interessant erweist sich etwa die Frage, wie „Erfolg“ aus Perspektive der Schüler\*innen aussähe. Hier ließe sich an Klafkis (1994) Vorschlag anschließen, Kriterien für Erfolg – hier die Erlangung „interkultureller und interreligiöser Kompetenz“ – gemeinsam zu erarbeiten und als Teil des Bildungsprozesses zu begreifen. Auch ließe sich ab einer bestimmten Altersstufe über den Einsatz von Peer-Researcher\*innen nachdenken, so dass Jugendliche selbst andere Jugendliche befragen. Durch die aktive Mitwirkung und Mitgestaltung würden Adressat\*innen der Evaluation gestärkt, wie es Roth (2004) fordert, und in ihren Sichtweisen ernst genommen. Partizipative Evaluationen können somit auch für Schüler\*innen einen empowernden Nutzen haben.

Enden wollen wir mit einem zweifachen Plädoyer, das sich aus unseren Erfahrungen ableitet: Zum einen müssen evaluierende Wissenschaftler\*innen verstärkt mit Auftraggeber\*innen, aber auch Projektverantwortlichen, einen Dialog führen über das zugrundeliegende Wirkungsverständnis. Im Sinne der vorgetragenen Erkenntnisse und Auffassungen halten wir es für wichtig, eine naive Input-Output-Orientierung in Frage zu stellen. Die Perspektive von Evaluationen sozialer, auch interkultureller Bildungsprozesse kann nicht die Evidenz im Sinne einer quasi algorithmischen „Punkt-für-Punkt-Abrechnung“ sein; denn, so Terhart, Pädagogik verfügt zwar über „prozesseigene Gütekriterien“, nicht aber über eine – entsprechend des Technologiedefizits in der Sozialen Arbeit – „deterministische Technologie“ (Terhart 2006:587).

Dazu gehört auch, dass die Bedingungen und Ressourcen für Forschung insgesamt ausgebaut und aufgestockt werden müssen, denn nur so lassen sich Wirkungen und Wirkfaktoren unter Einbeziehung aller Stakeholder



angemessen rekonstruieren. Partizipative und peer-gestützte Verfahren, teilnehmende Beobachtung, qualitative Befragungen und deren Auswertung etc. kosten in der Regel mehr Zeit (und Geld).

Zum anderen möchten wir dazu anregen, den „Trainingsbegriff“ aus non-formalen Bildungsangeboten wie dem hier im Mittelpunkt stehenden zu verabschieden. Die Suggestion, ein „sozialer Muskel“ oder eine „moralische Sehne“ seien trainierbar wie im Sport, bei dem es um klare, messbare Leistungsziele geht: Schnelligkeit, Treffsicherheit, Kraft etc., sollte erst gar nicht aufkommen. Form und Inhalt sind im Begriff „interkultureller Trainings“ nicht kongruent, denn: Der Begriff des Trainings steht im Widerspruch zu einem kritischen, nicht-funktionalen Bildungsbegriff – interkulturelle und interreligiöse „Lern-/Bildungsangebote“ jedoch zielen eben nicht auf Effektivität und Effizienz, sondern inhaltlich auf recht Gegenteiliges: kritische Reflexion, diskursive Aushandlung und Anstöße für Nach-Denklichkeit als Beitrag zu gesellschaftlicher Veränderung, Abbau von Vorurteilen und Diskriminierung und sozialer Gerechtigkeit.

## 7. Literatur

- Albus, S. / Polutta, A. (2008): Ergebnisse und Wirkungen im Feld der Sozialen Arbeit. In: Bielefelder Arbeitsgruppe (Hrsg.): *Soziale Arbeit in Gesellschaft*. Wiesbaden: VS-Verlag für Sozialwissenschaften, S 260-267.
- Böttcher, W. / Nüsken, D. (2015): Wirkungsforschung in der Kinder- und Jugendhilfe. In: *Jugendhilfe* 5, S. 348-355.
- Brohm, M. (2009): *Sozialkompetenz und Schule: Theoretische Grundlagen und empirische Befunde zu Gelingensbedingungen sozialbezogener Interventionen*. Weinheim und München: Juventa.
- Dollase, R. (2007). Kulturkonflikte oder ingroup-outgroup-Probleme?

Zu den schulischen Einflüssen auf die „interkulturelle“ Entwicklung von Kindern und Jugendlichen. In: Deutsche UNESCO-Kommission e.V. (Hrsg.): *Migration als Herausforderung. Praxisbeispiele aus den UNESCO-Projektschulen*. Bonn: Deutsche UNESCO-Kommission, S. 197-227.

- Galuske, M. (2013): *Methoden der Sozialen Arbeit*. 10. Auflage. Weinheim und Basel: Beltz Juventa.
- Haubrich, K. (2009). Evaluation in der Sozialen Arbeit in Deutschland: Entwicklungslinien und Besonderheiten der Evaluationsdebatte am Beispiel der Kinder-, Jugend- und Familienhilfe. In: Widmer, T. / Beywl, W. / Fabian, C. (Hrsg.): *Evaluation. Ein systematisches Handbuch*. Wiesbaden: VS-Verlag für Sozialwissenschaften, S. 442-449.
- Kanning, U. P. (2003): *Diagnostik sozialer Kompetenzen*. Kompendien. Psychologische Diagnostik. Bd.: 4. Göttingen: Hogrefe.
- Klafki, W. (1994): *Neue Studien zur Bildungstheorie und Didaktik: Zeitgemäße Allgemeinbildung und kritisch-konstruktive Didaktik*. 4. Auflage. Weinheim und Basel: Beltz.
- Lindner, W. (2015): Wirkfaktoren und Wirkungen in der Jugendarbeit. In: *Jugendhilfe* 5, S. 381-385.
- Müller-Kohlenberg, H. (2006): Evaluation und Selbstevaluation in der Sozialen Arbeit. In: Galuske, M. / Thole, W. (Hrsg.): *Vom Fall zum Management: Neue Methoden der Sozialen Arbeit*. Wiesbaden: VS Verlag für Sozialwissenschaften, S. 71-81.
- Otto, H.-U. / Polutta, A. / Ziegler, H. (Hrsg.) (2010): *What Works – Welches Wissen braucht die Soziale Arbeit. Zum Konzept evidenzbasierter Praxis*. Opladen und Farmington Hills: Barbara Budrich Verlag.
- Regiestelle Vielfalt gsub – Gesellschaft für soziale Unternehmensberatung mbH & Zentralstelle kompetent. für Demokratie - Stiftung Demokratische Jugend (Hrsg.) (o.J.): Abschlussbericht der Bundesprogramme „VIELFALT TUT GUT. Jugend für Vielfalt, Toleranz und Demokratie“ und „kompetent.

für Demokratie – Beratungsnetzwerke gegen Rechtsextremismus“ Förderphase 2007 – 2010. Berlin. Online Publikation. [https://www.demokratie-leben.de/fileadmin/content/PDF-DOC-XLS/Abschlussberichte/Abschlussbericht\\_barrierefrei.pdf](https://www.demokratie-leben.de/fileadmin/content/PDF-DOC-XLS/Abschlussberichte/Abschlussbericht_barrierefrei.pdf) (20.03.2019)

Roth, R. (2004): „Reden Sie mit dem Pferd.“ Bedingungen, Möglichkeiten und Grenzen Demokratischer Evaluationskultur in Deutschland. In: *sozial extra* 6, S. 6-9. DOI:10.1007/s12054-004-0065-8

Saldern, M. von (2001): *Ausländerkinder in Kleinstädten mit unterschiedlich hohen Ausländeranteil an der Wohnbevölkerung: Auswirkungen auf Einstellungen, soziale Motive und individuell-subjektive Wahrnehmung der schulischen Lern Umwelt*. Online Publikation. o.O. <http://opus.uni-lueneburg.de/opus/volltexte/2001/108/> [Zugriff am 30.09.2015]

Soellner, R. (2010): Modelle der Evaluation. In: Holling, H. / Schmitz, B. (Hrsg.): *Handbuch Statistik, Methoden und Evaluation*. Göttingen: Hogrefe, S. 223-260.

Stockmann, R. / Meyer, W. (2010): *Evaluation. Eine Einführung*. Opladen und Farmington Hills: Barbara Budrich Verlag.

Terhart, E. (2013): *Erziehungswissenschaft und Lehrerbildung*. Münster: Waxmann Verlag.

Terhart, E. (2006): Professionalität im Lehrerberuf. Ratlosigkeit der Theorie, gelingende Praxis. In: *Zeitschrift für Erziehungswissenschaft* 4, S. 580-597. DOI: 10.1007/s11618-006-0169-y

Wahl, D. (2002): Mit Training vom trägen Wissen zum kompetenten Handeln? In: *Zeitschrift für Pädagogik* 2, S. 227-241.

Zweifel an der Professionalität pädagogischer und sozialarbeiterischer Maßnahmen gelesen werden: „Die Legitimationsfrage stellt sich im Rahmen der Wirkungsdebatte besonders deutlich“ (Albus/Polutta 2008:260).

2 Stockmann und Meyer (2010:73) benennen vier Funktionen von Evaluationen: Erkenntnis, Kontrolle, Legitimation und Entwicklung.

3 Die multiethnische und multireligiöse Teamzusammensetzung gehört zum festen Bestandteil des Programms.

4 Weiterentwicklungen in Richtung einer „Empowerment-Evaluation“ nach Fetterman mit dem Ziel „to help people themselves and improve their programs using a form of self-evaluation and reflecting“ (Fetterman 2000, zit. n. Müller-Kohlenberg 2006:79), wären zu diskutieren. Ziel dieser Form der Empowerment-Evaluation ist es, v.a. die Mitarbeiter\*innen der zu evaluierenden Einrichtungen so zu qualifizieren und zu stärken, dass sie sich selbst evaluieren können, wobei Parteilichkeit ausdrücklich Bestandteil der Evaluation ist. Selbstevaluationen werden damit zur Methode der Wahl, die Fremdevaluationen überflüssig werden lassen. Das Für und Wider von Selbst- versus Fremdevaluation kann an dieser Stelle jedoch nicht weiter ausgeführt werden (weiterführend z.B. Müller-Kohlenberg 2006).

## 8. Endnotes

1 Die neuerliche Debatte um Ergebnisse und Wirkungen pädagogischer und sozialer Praxis und der Ruf nach Wirkungsforschung und Evidenzbasierung kann auch als „Effektivitätskritik“ (Albus/Polutta 2008:261), d.h. als grundsätzlicher

