

Jugendgruppenfahrten im Spiegel mehrebenenanalytischer Untersuchungen

Erfahrungen mit vernetzter Selbstevaluation in
non-formalen Bildungssettings

*Wolfgang Ilg, Michael Diehl*¹

Zusammenfassung: In den Jahren 2005 bis 2010 wurden bei 679 Freizeiten und internationalen Jugendbegegnungen quantitative Daten erhoben. Zum Einsatz kamen ein Fragebogen für Teilnehmende, ein Ziele-Fragebogen für Mitarbeitende sowie ein Fragebogen zur Erfassung von Rahmenbedingungen der Maßnahme. Mithilfe der Mehrebenenanalyse wird ein Teildatensatz (5.136 Teilnehmende, 973 Mitarbeitende) im Blick auf die Kriterien Persönlichkeitsentwicklung, politische Reflexion und interkulturelle Erfahrung ausgewertet. Deutliche Unterschiede auf Gruppenebene ergeben sich zwischen Freizeiten und Jugendbegegnungen. Die Ziele der Mitarbeitenden erweisen sich als gute Indikatoren für das pädagogische Konzept und tragen zur Varianzaufklärung erheblich bei. Die hier vorgestellte Methode der vernetzten Selbstevaluation ermöglicht auf einfache Weise, im non-formalen Bildungsbereich Evaluationsdaten sowohl für die lokale Auswertung als auch für übergreifende Analysen zu generieren.

Schlagwörter: Freizeiten, Jugendbegegnungen, Mehrebenenanalyse, Selbstevaluation

Youth camps and international encounters in the light of multi-level analysis

A model of joint self-evaluation in non-formal education

Abstract: The present article is a report about a quantitative study based on 679 youth camps and international youth encounters in Europe which took place between 2005 and 2010. The questionnaires addressed the participants (asking for their experiences), the workers (asking for their aims) and the leaders (asking for basic data). Selected data (5.136 participants, 973 workers) is analysed by multi-level analysis for identifying predictors concerning three criteria variables: personality, political reflexion and intercultural experiences. The analysis shows substantial differences in the results of camps versus international encounters. The aims of the workers are used as indicators for the pedagogical concept and contribute to a considerable degree to the explanation of variance. The method of joint self-evaluation presented here shows how evaluation data can be generated in non-formal educational settings, thus producing data for local use as well as for the comprehensive analysis on a general level.

Keywords: Camps, Youth Encounter, Multi-Level Analysis, Self-Evaluation

¹ Beide: Psychologisches Institut, Abteilung Sozial- und Persönlichkeitspsychologie, Universität Tübingen

1. Einleitung

Anekdotisch sind sie vielfach belegt – die Bildungsleistungen von Reisen in einer Jugendgruppe. Etliche Menschen erinnern sich auch nach Jahrzehnten noch an die Gemeinschaftserfahrungen beim Zeltlager der Jugendfeuerwehr oder an die zweisprachigen Theaterproben beim Kulturprojekt ihrer deutsch-polnischen Jugendbegegnung. Die Übernahme von Verantwortung in der Gruppe, das Verständnis für politische Zusammenhänge oder die Beschäftigung mit der Kultur des eigenen und eines fremden Landes – solche und andere Lernerfahrungen geschehen bei einer gemeinsamen Reise oft unmittelbarer und intensiver als in der Schule. Über die ‚anekdotische Evidenz‘ hinaus sind empirische Belege für solche Bildungswirkungen jedoch – wie in der Jugendarbeit insgesamt – rar. Angesichts der gewachsenen Bedeutung empirischer Bildungsforschung im schulischen Bereich kommt aber auch die außerschulische Bildung nicht umhin, ihre Arbeit zu evaluieren. Sind die von den Verantwortlichen behaupteten Bildungsleistungen tatsächlich vorhanden? Inwiefern lassen diese sich pädagogisch steuern und spezifisch anbahnen? Ist das, was ein Jugendlicher bei einer solchen Fahrt erlebt, vor allem davon abhängig, welche individuellen Voraussetzungen sein Erleben prägen oder spielt das von den Mitarbeitenden verantwortete pädagogische Konzept eine nachweisbare Rolle?

Der vorliegende Artikel zeigt auf, welche Möglichkeiten und Grenzen mit der empirischen Erfassung non-formaler Bildungsleistungen verbunden sind. Die mehrbenenanalytischen Auswertungen stützen sich dabei auf zwei miteinander verzahnte Forschungsprojekte, in denen seit 2005 erstmals standardisierte Evaluationsverfahren für Jugendfreizeiten sowie internationale Jugendbegegnungen entwickelt und implementiert wurden. Ausgehend von je einer Grundlagenstudie für Freizeiten und Begegnungen wurde für beide Arbeitsformen ein einheitliches Verfahren entwickelt, mit dem Träger vor Ort ihre Fahrt selbständig evaluieren können und diese Daten dann zu einer zentralen Datensammelstelle einsenden. Im Folgenden wird dieses Verfahren als ‚vernetzte Selbstevaluation‘ bezeichnet. Neben der inhaltlichen Auswertung der Daten werden die ersten vier Jahre dieser vernetzten Selbstevaluation auch im Hinblick auf die entwickelte Evaluationsmethodik reflektiert, um mögliche Anwendungsgebiete eines solchen Verfahrens für weitere non-formale Bildungsbereiche auszuloten.

1.1 Evaluation non-formaler Bildungsfelder

Als Zwischenform zwischen formaler Bildung (Schule und Berufsbildung) und informeller Bildung (die alle bildenden Aspekte des Alltagsgeschehens umfasst), bezieht sich der non-formale Bildungsbereich auf Formen institutionalisierter Bildung, die zwar freiwillig und ohne Zertifizierungen verlaufen, aber dennoch klaren Bildungs-Konzeptionen folgen (vgl. Bundesjugendkuratorium 2001; BMFSFJ 2005). In den bildungspolitischen Diskussionen der letzten Jahre werden dem non-formalen Bereich, insbesondere den verschiedenen Formen der Jugendarbeit, Bil-

dungsleistungen zugeschrieben, die zwar nicht direkt zur formalen Bildung beitragen, aber die Voraussetzungen für den Schul- und Ausbildungserfolg oftmals befördern, indem personale und soziale Kompetenzen wie Teamfähigkeit oder der Umgang mit Konflikten anhand persönlich relevanter Situationen eingeübt werden (vgl. Rauschenbach 2009).

Die Evaluation der non-formalen Bildung hat mit der Evaluation im schulischen Bereich in den letzten Jahren allerdings nicht annähernd Schritt gehalten. Noch immer sind die Vorbehalte bei Praktiker(inne)n der Jugendarbeit gegenüber empirischen Studien groß. So spricht Zitzmann (2007) von einer „fortschreitende[n] Entfremdung von Wissenschaft und Praxis“ in der Jugendarbeit. Aus Sicht der Wissenschaftler(innen) konstatiert sie ein „Evaluierungsvakuum“ und einen Mangel an „systematische[n] Evaluationen der potentiellen Wirkungen pädagogischer Maßnahmen“. Aus Sicht der Praktiker(innen) führt sie als häufig vorgebrachtes Gegenargument an: „Auch mit feinsten Instrumenten kann die entscheidende Wirkung unserer Arbeit nicht gemessen und belegt werden“ (vgl. auch Sturzenhecker/von Spiegel 2009). Überblicksarbeiten, die im Umfeld des Deutschen Jugendinstituts (DJI) in den letzten Jahren entstanden, verdeutlichen zwar das zunehmende Interesse an empirischer Evaluationsforschung im Bereich der Jugendarbeit, können aber fast durchweg nur regional begrenzte Evaluationen mit geringen Fallzahlen benennen (vgl. Arbeitskreis G5 2009, Arbeitskreis G5 2010; Rauschenbach et al. 2010). Grundlegende theoretische Vorarbeiten für Evaluationen im Bereich der Jugendarbeit wurden im DJI insbesondere im Projekt „Strategien und Konzepte externer Evaluation in der Kinder- und Jugendhilfe – eXe“ vorgelegt. Lüders und Haubrich (2006) verorten dabei die Wirkungsfrage als eine Grundfrage pädagogischen Handelns, weisen aber auf die notwendigen Differenzierungen im Blick auf die häufig zu hörende Rede von den ‚Wirkungen‘ in der Jugendarbeit hin. Für die Beobachtung von Wirkungen im Kontext der Jugendhilfe folgern sie, dass Entscheidungen darüber getroffen werden müssen, welche Ziele und welche Wirkungen beobachtet werden sollen. Damit treffen die Beteiligten eine Entscheidung darüber, welcher „Realitätsausschnitt im Sinne eines Beobachtungsausschnittes“ als relevant angesehen wird und welche anderen Fragestellungen notwendigerweise unbeobachtet bleiben. Beywl benennt als besondere Herausforderungen der Wirkungsmessung in der Kinder- und Jugendarbeit die Koproduktion des Bildungsgeschehens durch Mitarbeitende und Jugendliche sowie die sich aus der Trägerpluralität ergebende Wertebestimmtheit der jeweiligen Jugendverbände. Der von ihm entwickelte „Programmbaum“ versinnbildlicht, dass auch non-formale Bildungsarbeit von den erwünschten „Früchten“ am Baum her gedacht werden muss: „Das gesamte Programm wird von den *outcomes* her geplant“, alle Aktivitäten hängen „von ihrer (optimalerweise durch Jugendforschung) begründeten Fähigkeit ab, gewünschte *outcomes* auszulösen“ (Beywl 2006: 38).

In der Praxis der Jugendarbeit konnten sich bislang noch kaum Evaluationsverfahren etablieren, die sowohl den für Wissenschaftler(innen) bedeutsamen Gütekriterien von Reliabilität und Validität als auch den von Praktiker(inne)n eingeforderten Bedingungen von Praktikabilität, Nützlichkeit und Ökonomie genügen. Die im Folgenden vorgestellte vernetzte Selbstevaluation stellt eines der wenigen

Evaluationsverfahren dar, das bereits auf eine erste Implementierungsphase in der Praxis der Jugendarbeit zurückblicken kann. Das zu evaluierende Programm, die Gruppenfahrten im Kontext der Jugendarbeit, sollen zunächst näher beschrieben werden.

Unter einer *Jugendgruppenfahrt* wird eine mit Gruppen durchgeführte, freiwillige, nicht am Heimatort stattfindende Aktivität verstanden, die mehr als zwei Tage dauert und deren Zielsetzung über die bloße Organisation eines gemeinsamen Urlaubs hinaus pädagogisch begründet und von Erwachsenen begleitet wird (vgl. Ilg 2005: 21). Trotz einer Vielzahl denkbarer Gestaltungsformen können zwei Grundtypen solcher Fahrten unterschieden werden: Als *Freizeit* wird eine Maßnahme verstanden, bei der eine Gruppe aus einem Land (hier: Deutschland) gemeinsam mit einem Mitarbeiter(innen)-Team (zumeist im Ausland) unterwegs ist. Der zweite Typus ist die *internationale Jugendbegegnung*, die sich dadurch auszeichnet, dass die Gruppe nicht unter sich bleibt, sondern sich mit ein, zwei oder mehreren ähnlichen Gruppen aus anderen Ländern trifft, also eine bi-, tri- oder multilaterale Begegnung durchführt. Im Hinblick auf die äußeren Rahmenbedingungen sind die Unterschiede zwischen Freizeiten und internationalen Jugendbegegnungen vergleichbar. Beide Typen von Gruppenfahrten werden zumeist von freien Trägern der Jugendarbeit angeboten (z.B. Sportverein, kirchlicher Jugendverband, Jugendzirkus), sprechen schwerpunktmäßig Jugendliche im Teenageralter an, finden in der Regel in den Schulferien statt und dauern ungefähr zwei Wochen. Begleitet werden die Jugendlichen jeweils von einem Team aus haupt- und ehrenamtlichen Mitarbeitenden, die das Programm vorbereiten (zum Überblick vgl. Müller, o.J.). Im Blick auf die pädagogische Programmatik stehen bei internationalen Jugendbegegnungen die Aspekte politischer und interkultureller Bildung sehr viel stärker im Vordergrund als das bei Freizeiten der Fall ist – was unter anderem die finanziell sehr viel stärkere Förderung der Begegnungen durch öffentliche Mittel begründet.

1.2 Datengewinnung durch vernetzte Selbstevaluation

Bei Jugendgruppenfahrten kommt der Auswertung des gemeinsam Erlebten eine wichtige Bedeutung zu. Ein Programm, das sich – auch von seinem gesetzlichen Auftrag in Paragraf 11 des Kinder- und Jugendhilfegesetzes – die Mitgestaltung durch die Jugendlichen zu einem wichtigen Leitprinzip gemacht hat, kann kaum umhin, die Jugendlichen in die Bewertung der Programme einzubeziehen. Die bislang eingesetzten Feedbackmethoden verwenden jedoch großteils ein spielerisches Format, das nicht auf systematische Quantifizierungen der Rückmeldungen oder deren schriftliche Dokumentation abzielt.²

Von den bislang ad hoc entwickelten Evaluationsmethoden bei Jugendgruppenfahrten zu einer standardisierten Form der Datengewinnung zu gelangen, war und ist das Ziel des Projekts ‚Freizeitenevaluation‘. Seit dem Jahr 2001 wurden in die-

2 Zum Überblick vgl. Kloosterman/Giebel/Senyuva 2007, für das Deutsch-Französische Jugendwerk vgl. Müller 1996, Ansätze zu einer quantitativen Selbstevaluation in den USA beschreiben Garst/Bruce 2003.

sem Projekt verschiedene Evaluationsverfahren entwickelt, die es den Veranstaltern von Freizeiten und internationalen Jugendbegegnungen ermöglichen, ihre Jugendgruppenfahrten durch ein lokal durchgeführtes quantitatives Verfahren selbständig zu evaluieren.³ In bereichsspezifischen Grundlagenstudien (vgl. Ilg 2008a für Freizeiten, Dubiski/Ilg 2008 für internationale Jugendbegegnungen des Deutsch-Französischen Jugendwerks und des Deutsch-Polnischen Jugendwerks) wurden Fragebögen entwickelt, die nun in einer optimierten Kurzfassung für die eigenständige Evaluation zur Verfügung stehen. Veranstalter von Jugendgruppenfahrten können die existierenden Fragebögen durch eigene Items ergänzen und erhalten nach der Dateneingabe deskriptive Auswertungen am eigenen Computerbildschirm. Da die Entwicklung der Verfahren durch Mittel des Bundesministeriums für Familie, Senioren, Frauen und Jugend (BMFSFJ) finanziert und die einfache Statistik-Software ‚GrafStat‘ von der Bundeszentrale für politische Bildung zur Verfügung gestellt wurde, bleibt die Nutzung der Evaluationsverfahren für die Veranstalter kostenfrei.

Während die Grundlagenstudien zur Sicherung von Reliabilität und Validität neben der Befragung direkt bei der Fahrt auch eine Nachbefragung der Teilnehmenden drei Monate später einsetzten und zahlreiche Rahmendaten u.a. zur Programmstruktur erhoben, bleiben die Materialien zur Selbstevaluation auf drei Bausteine begrenzt: Von zentraler Bedeutung ist der Fragebogen für die Teilnehmenden, den diese am vorletzten Tag der Fahrt ausfüllen und der die (subjektiv empfundenen) ‚outcomes‘ erfasst. In einem weiteren Fragebogen zur Erfassung der Ziele werden vor Beginn (bzw. direkt zu Beginn) der Maßnahme die Intentionen der Mitarbeitenden erhoben. Ein dritter kurzer Fragebogen schließlich erfasst wichtige Rahmendaten wie die Dauer der Maßnahme oder die Gruppengröße. Dieser Fragebogen wird von der verantwortlichen Leitungsperson ausgefüllt. Das Verfahren zielt in erster Linie darauf ab, Veranstaltern vor Ort Instrumente zur einfachen und verlässlichen Selbstevaluation zur Verfügung zu stellen. Darüber hinaus werden die Veranstalter gebeten, die lokal erhobenen Daten an eine zentrale Stelle einzusenden, um so eine Gesamtauswertung auf überregionaler Ebene zu ermöglichen. Als Anreiz für das Einsenden der Daten erhalten die Veranstalter eine sogenannte ‚Kompaktauswertung‘, also eine tabellarische und grafische Übersicht ihrer Daten, die von GrafStat nicht in dieser Form zur Verfügung gestellt wird. Allerdings kann keinerlei Druck auf die örtlichen Veranstalter zur Einsendung der Daten ausgeübt werden. Die Motivation der Veranstalter zur Dateneinsendung und die Entwicklung der Einsendezahlen gehören daher zu den Fragen, die in der ersten Implementierungsphase erkundet werden sollten.

3 Für einen Überblick über die Projektentwicklung vgl. Ilg 2008b; zur Begründung der Methodik vgl. Ilg 2010; vgl. insgesamt <http://www.freizeitenevaluation.de>.

1.3 Stand der Forschung

Wie für den non-formalen Bildungsbereich insgesamt, liegen auch für Jugendgruppenfahrten in Deutschland nur wenige empirische Studien vor: So stammen im Forschungsüberblick „Das Wissen zur Kinder- und Jugendarbeit“ (Arbeitskreis G5 2009) im Bereich der Jugendfreizeiten alle drei vorgestellten Studien aus dem Kontext der Freizeitevaluation. Für die internationalen Jugendbegegnungen sind zwar Langzeitwirkungen, vornehmlich auf der Basis von Interviewdaten, erforscht (vgl. Thomas/Chang/Abt 2007), Befragungen von aktuellen Teilnehmer(inne)n von Begegnungen fehlen aber weitgehend (von Studien zu spezifischen Programmen mit begrenztem Datenumfang abgesehen, so zum Beispiel die Evaluation des Jugendbegegnungsprojekts „Eight Columns“; vgl. MGFFI 2007). Auch international wird dem Arbeitsfeld der Jugendgruppenfahrten bislang wenig Aufmerksamkeit gewidmet – mit Ausnahme einer Forschergruppe in den USA.

Dort wurde – zeitgleich mit der Entwicklung des hier vorgestellten Evaluationsverfahrens – in den letzten Jahren eine umfassende Feldforschung im Auftrag der American Camp Association durchgeführt (ACA, vgl. <http://www.acacamps.org/research>). Die US-Forschergruppe (Thurber et al. 2007; Henderson et al. 2006a) etablierte in den Jahren 2002 und 2003 eine groß angelegte Studie zu Camps in den USA, die trotz mancher Unterschiede etliche Parallelen zu den hier untersuchten Jugendgruppenfahrten aufweisen. Diese Studie konnte in fast allen erfragten Bereichen einen kleinen, aber signifikanten Anstieg der Selbst- und Fremdeinschätzungen zwischen den pre-camp und post-camp Ergebnissen nachweisen, im follow-up blieben die höheren Werte zumeist erhalten. Im Blick auf mögliche Prädiktoren der Effekte wurden verschiedene Variablen untersucht: Die Länge des Camp-Aufenthalts zeigte keinen signifikanten Einfluss. Auch Geschlecht und ethnische Herkunft spielten keine Rolle. Dagegen konnte gezeigt werden, dass Camps von religiösen Organisationen stärkere Effekte im Bereich „spiritual growth“ zeigten als andere Camps.

Die US-Forscher(innen) resümieren, dass mit ihrer Studie zwar ein erster wichtiger Schritt getan worden sei, der aber zugleich weitere Forschungsnotwendigkeiten verdeutliche: Insbesondere fehle es an einem Vergleich, inwiefern die 80 Camps zu jeweils eigenen Resultaten führen: „This study did not permit valid comparisons among the 80 participating camps. Had such comparisons been possible, we might have learned more about what camp programs, missions, or philosophies were associated with the most growth. There is more to learn about intentionality in camp programming“ (Thurber et al. 2007: 252). Der Versuch zur Erfassung des pädagogischen Konzepts über die Ziele der „camp directors“ (Henderson et al. 2006b) erwies sich als wenig ergiebig. Lediglich in zwei Bereichen, *physical and cognitive skills* sowie *spiritual growth*, zeigten sich relevante Varianzanteile, die mit den Zielsetzungen der camp directors im Zusammenhang stehen. Die Forscher(innen) konstatieren hier weiteren Forschungsbedarf: „The lack of evidence about the relationship between intended goals and youth development may be due to the global nature of the goals presented to the directors. More research is needed to understand the relationship between youth development and camp programs, structures, and processes“

(Henderson et al. 2006b: 8). Ebenso gelang es nicht, Prädiktoren für die unterschiedlichen Bildungsprozesse bei den Camps aus einer Analyse der jeweiligen Programmkomponenten zu gewinnen (Henderson et al. 2007).

1.4 Das pädagogische Konzept als Prädiktor des Gruppengeschehens

Der von Henderson et al. (2006b) beklagte Mangel einer adäquaten Operationalisierung des pädagogischen Konzepts ist zugleich eine Stärke der (unabhängig von den US-Forschungen entwickelten) hier dargestellten Evaluationsstudie zu Freizeiten und internationalen Jugendbegegnungen in Deutschland und Europa. Ein Ziel dieses Evaluationsverfahrens ist es, Indikatoren für das pädagogische Konzept eines Leitungsteams zu identifizieren, die sich in der Praxis gut erheben lassen und hinsichtlich ihrer prädiktiven Aussagekraft untersucht werden können. Anders als im Ansatz der US-Forscherguppe werden hierfür nicht Programmkomponenten und auch nicht die Ziele lediglich einer bzw. eines Hauptverantwortlichen, sondern die Zielsetzungen aller Mitarbeitenden als Prädiktoren verwendet.

Im Einklang mit den ernüchternden Ergebnissen von Henderson et al. (2007) wird davon ausgegangen, dass sich die pädagogische Praxis einer Jugendgruppenfahrt nicht über die Erhebung von Programmstrukturen adäquat beschreiben lässt. Relevante pädagogische Vorgänge ereignen sich im 24-Stunden-Rahmen von Jugendgruppenfahrten oftmals außerhalb des eigentlichen Programms, beispielsweise, wenn Mitarbeitende auf ein unvorhergesehenes Ereignis reagieren müssen und dabei auf ihren ‚inneren Kompass‘, also die eigenen Ziele und ihr handlungsorientiertes Wissen, angewiesen sind. Die Zielbeschreibungen von Mitarbeitenden wurden daher als Indikator für die pädagogische Programmatik verwendet. In dem für die Freizeitevaluation entwickelten Evaluationsdesign (ausführlich vgl. Ilg 2008a: 25-31) gewichten die Mitarbeitenden vor bzw. zu Beginn der Fahrt verschiedene vorgegebene Zielsetzungen (z.B.: „Die Teilnehmer⁴ sollen die Kultur der Gastregion kennen lernen“). Analog zu diesen subjektiven Zielsetzungen, die als Indikatoren pädagogischer Konzeption verwendet werden, bezieht sich die Messung der *outcomes* nicht auf (faktisch kaum objektivierbare) Erlebnisse, sondern auf die subjektiven Wahrnehmungen der Jugendlichen am Ende der Maßnahme. Abbildung 1 stellt den Zusammenhang grafisch dar und illustriert ihn anhand eines Beispiels.

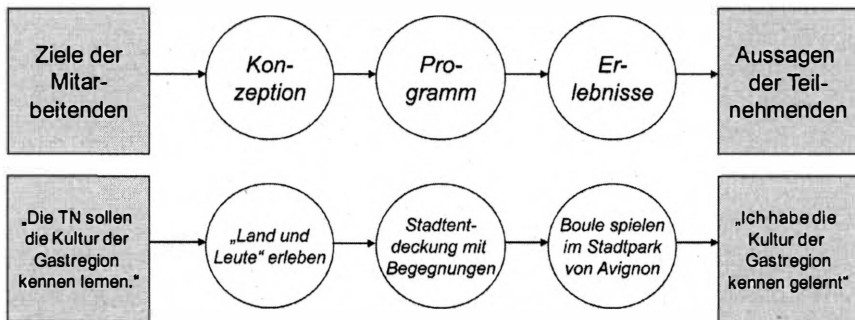
Der neue Ansatz gegenüber dem von Henderson et al. (2006b) eingesetzten Verfahren besteht darin, dass nicht nur die Hauptzuständigen für eine Fahrt nach ihren Zielen befragt werden, sondern alle Mitarbeitenden, die bei der entsprechenden Maßnahme eine pädagogische Aufgabe haben. In den Kontexten der Jugendarbeit herrscht zwischen Hauptleitung und Team zumeist ein gleichrangiges Mit-

4 Zugunsten einer besseren Lesbarkeit wurde in den Fragebögen zur Bezeichnung beider Geschlechter jeweils nur die männliche Schreibweise verwendet. Dies wird auch im Folgenden bei Frageformulierungen sowie der Bezeichnung der beiden Stichproben/Instrumente beibehalten. Selbstverständlich sind mit „Mitarbeitern“ und „Teilnehmern“ jedoch stets auch Mitarbeiterinnen und Teilnehmerinnen gemeint.

einander, so dass die Fokussierung auf die Leitungsperson keine ausreichende Auskunft zu den real vorhandenen Zielvorstellungen der pädagogisch Tätigen bietet.

Mit der Frage nach den von den Mitarbeitenden verfolgten Zielen wird die Definition von Erfolgskriterien den jeweiligen Jugendverbänden überlassen. Dahinter steht die Einsicht, dass im non-formalen Bildungsbereich – anders als in der Schule – keine zentral definierten Erfolgskriterien existieren. Vielmehr bestimmen die Akteure der jeweiligen Bildungsmaßnahme innerhalb eines vorgegebenen Rahmens die von ihnen favorisierten Ziele mit. Als erfolgreich im Sinne des Veranstalters kann eine Jugendgruppenfahrt dann gelten, wenn die von den dafür verantwortlichen Mitarbeitenden selbst gesetzten Ziele erreicht werden (unter der Voraussetzung, dass diese im Einklang mit den formulierten Absichten des Veranstalters stehen – dieser Aspekt wird hier nicht näher untersucht).

Abbildung 1: Zusammenhang von Mitarbeiter-Zielen und Teilnehmer-Aussagen: allgemein (oben) und an einem Beispiel (unten)



Anmerkungen: Die in Kreisen dargestellten Aspekte lassen sich empirisch kaum erfassen. Als Indikatoren für das pädagogische Konzept werden daher Ziele der Mitarbeitenden erhoben, als Indikatoren für das Erleben der Freizeit die Aussagen der Teilnehmenden. Vgl. zur Grundidee den Programmbaum von Beywl 2006.

Aus der Vielzahl der denkbaren Zielsetzungen wurden für die vorliegende Analyse drei Kriterien ausgewählt: Persönlichkeitsentwicklung, politische Reflexion und interkulturelle Erfahrung. Der Begriff der Persönlichkeitsentwicklung greift auf die Grundlagenarbeit von Nagl (2000) zurück. Nagl beschreibt mit dem Begriff der „Personalisation“ ein wesentliches Ziel pädagogischer Gruppenarbeit: Jugendliche sollen demzufolge die in ihnen liegenden Möglichkeiten entdecken und spielerisch ausprobieren, um Orientierungen für die Ausbildung einer eigenständigen Persönlichkeit zu erhalten. Das Kriterium der Persönlichkeitsentwicklung stellt also ein für Aktivitäten der Jugendarbeit allgemein angenommenes Kriterium dar und sollte sich daher sowohl bei Freizeiten als auch bei internationalen Jugendbegegnungen ergeben. Die Zielsetzung politischer Reflexion wurde ausgewählt, weil es sich hier um ein in der öffentlichen Diskussion vielfach gefordertes Ziel von Jugendarbeit handelt. Während der Urlaubs-Charakter von Jugendfreizeiten die Beschäftigung mit politischen Inhalten eher erschweren dürfte, ist zu erwarten, dass sich bei inter-

nationalen Jugendbegegnungen dieses Thema leichter integrieren lässt. Mit dem Begriff der interkulturellen Erfahrung greift das dritte Kriterium einen Bereich auf, der zwar potenziell sowohl bei Freizeiten als auch bei internationalen Jugendbegegnungen relevant werden kann, von den Vertreter(inne)n der Begegnungsarbeit aber als ein Spezifikum internationaler Arbeit hervorgehoben wird.

1.5 Fragestellungen und Hypothesen

Die Fragestellungen beziehen sich zunächst auf die *Ergebnisse* der bereits durchgeführten Evaluationen von Freizeiten und internationalen Jugendbegegnungen:

1. Inwiefern unterscheidet sich das Erleben der Jugendlichen im Vergleich zwischen den einzelnen Gruppen?
2. Welche Prädiktoren auf individueller Ebene (Alter, Geschlecht, Vorerfahrung und Nationalität) und Gruppenebene (Typus der Gruppenfahrt, Gruppengröße) zeigen Effekte auf die Erfahrungen der Jugendlichen bei den Fahrten? Insbesondere die Unterschiede zwischen den Fahrt-Typen erscheinen von Interesse: Sind die interkulturellen Erfahrungen bei einer Jugendbegegnung höher als dies mit einem einfachen Auslandsaufenthalt einer deutschen Jugendgruppe erreicht wird?
3. Welcher zusätzliche Beitrag zur Varianzaufklärung ergibt sich, wenn die Ziele der Mitarbeitenden als konzeptionelle Prädiktoren in die Analyse mit eingehen?
4. Zeigt sich für die untersuchten Kriteriumsbereiche ein spezifischer Einfluss der Mitarbeitendenziele, so dass konzeptionelle Schwerpunktlegungen in einem bestimmten Bereich sich in den Rückmeldungen der Jugendlichen widerspiegeln? Erwartet wird, dass dieser spezifische Zusammenhang bei allen drei Kriteriumsbereichen besteht, da alle drei eher auf individuelle Vorgänge und nicht auf Prozesse der Gemeinschaftsbildung abzielen (vgl. Ilg/Lüdtke 2011).

In einem zweiten Schritt soll dann die *Evaluationsmethode* einer empirischen Überprüfung unterzogen werden. Hier sind folgende Fragestellungen von Interesse:

5. Welche Chancen und Grenzen ergeben sich für ein Konzept der vernetzten Selbstevaluation? Lässt sich bei einem lokal autonom einsetzbaren Evaluationssystem der Datenrücklauf zu einer zentralen Datensammelstelle gewährleisten? Wie kann die Motivation für die Dateneinsendung erhöht werden?
6. Sind die Daten, die in einer Selbstevaluation durch Fachkräfte ohne Evaluation Ausbildung erhoben werden, hinsichtlich der Gütekriterien für eine wissenschaftliche Auswertung verwendbar?

2. Methode

2.1 Stichprobe

Die vorliegende Analyse bezieht sich auf einen Teildatensatz der beiden Evaluationsprojekte, die im Jahr 2005 entwickelt wurden und seit 2006 (Freizeiten) bzw. 2007 (internationale Jugendbegegnungen) im vernetzten Selbstevaluationsverfahren kontinuierlich eingesetzt werden. Insgesamt liegen Daten von 679 Maßnahmen mit 1.952 Mitarbeitenden und 15.604 Teilnehmenden aus den Jahren 2005 bis 2010 vor. Bei den Maßnahmen handelt es sich jeweils ungefähr zur Hälfte um Freizeiten für Jugendliche aus Deutschland, zur anderen Hälfte um internationale Jugendbegegnungen. Zum Prinzip der vernetzten Selbstevaluation gehört die Freiheit der lokalen Jugendarbeits-Träger, zu entscheiden, ob und wie umfangreich sie die angebotenen Evaluationsinstrumente (Fragebogen für Teilnehmer, Mitarbeiter, Leitung) einsetzen. Viele Träger verwendeten nur die Fragebögen für Teilnehmer, nicht aber die Fragebögen für Mitarbeiter und Leitungen. Der hier vorgestellte Teildatensatz bezieht nur solche Maßnahmen ein, bei denen sowohl Teilnehmer- als auch Mitarbeiter-Fragebögen vorliegen. Die Fallzahlen sind in Tabelle 1 dargestellt.

Tabelle 1: Fallzahlen für die Mehrebenenanalyse

	<i>Maßnahmen</i>	<i>Mitarbeiter</i>	<i>Teilnehmer</i>
Freizeiten	96	360	2.100
Internationale Begegnungen	146	613	3.036
Summe	242	973	5.136

Anmerkungen: Die Fragen zur interkulturellen Erfahrung werden aufgrund einer Filterfrage nur von denjenigen Jugendlichen ausgefüllt, die sich im Ausland aufhalten. Die Anzahl gültiger Fälle bei den Teilnehmenden liegt daher für das Kriterium ‚interkulturelle Erfahrung‘ bei 3.695. Die mehrebenenanalytische Berechnung wurde im Jahr 2010 durchgeführt und bezieht sich daher nur auf die Daten von 2005-2009.

Im vorliegenden Datensatz gehören 64% der Teilnehmenden zu einer deutschen, 10% zu einer französischen, 8% zu einer polnischen und 17% zu einer Gruppe aus einem anderen Land. 59% der Teilnehmenden sind weiblich, das Durchschnittsalter liegt bei 16 Jahren. 41% der befragten Jugendlichen sind zum ersten Mal bei dieser Art von Jugendgruppenfahrt dabei.

2.2 Instrumente und Variablen

Für die Einzelverfahren der bundesweiten Freizeitevaluation und der Evaluation internationaler Jugendbegegnungen liegen mittlerweile verschiedene Skalen vor, von denen die meisten über eine gute interne Konsistenz verfügen (vgl. 4.2). Allerdings wurden diese Skalen spezifisch für den jeweiligen Typus der Jugendgruppenfahrt (Freizeit oder internationale Begegnung) entwickelt. Für die in der vorliegen-

den Untersuchung durchgeführte gemeinsame Analyse mussten Kriteriumsvariablen gefunden werden, die in beiden Fahrt-Typen verwendet werden konnten. Zwei der drei Kriterien lassen sich durch einen Index abbilden, der Bereich politischer Reflexion beruht dagegen lediglich auf einem Item. Sowohl die Ziele-Befragung bei den Mitarbeitenden als auch die Erfassung der Erfahrungen bei den Teilnehmenden erfolgte über eine 7-stufige Likert-Skala, die Werte wurden z-standardisiert. Daneben gehen Nominalvariablen ein, die jeweils mit 0 für „nicht zutreffend“ und 1 für „zutreffend“ kodiert wurden.

Die drei *Kriteriumsvariablen* aus dem Teilnehmer-Fragebogen sind: *Persönlichkeitsentwicklung* (Items: „Ich habe neue Seiten und Fähigkeiten bei mir entdeckt“; „Ich habe hier neue Interessen und Hobbys entdeckt“; „Es gab hier viele Möglichkeiten, mich kreativ zu betätigen“; $\alpha=.67$), *politische Reflexion* („Ich habe mich hier mit gesellschaftlichen/politischen Themen auseinandergesetzt“) und *interkulturelle Erfahrung* („Ich habe meine Fremdsprachenkenntnisse verbessert“; „Ich habe Alltag und Kultur der Gastregion kennen gelernt, z.B. Essen, Musik, Traditionen“; „Ich bin in Kontakt mit Menschen aus der Gastregion gekommen“; $\alpha=.72$).

Die *Prädiktoren auf Individualebene* beziehen sich auf Alter und Geschlecht der Teilnehmenden sowie auf deren Vorerfahrung mit der entsprechenden Art von Jugendgruppenfahrt. Als Jugendlicher ohne Vorerfahrung gilt, wer zum ersten Mal bei einer solchen Freizeit bzw. Jugendbegegnung teilnimmt. Zudem geht das Herkunftsland der Jugendlichen in die Analyse ein. Da der Anteil von in Deutschland lebenden Jugendlichen mit nicht deutscher Staatsangehörigkeit unter den Teilnehmenden von Freizeiten und internationalen Jugendbegegnungen nur bei rund 3% liegt (vgl. Ilg 2007, Dubiski 2010), geht der Faktor Migrationshintergrund nicht eigens in die Analyse ein.

Als *allgemeine Prädiktoren auf Gruppenebene* werden der Typ der Gruppenfahrt (Freizeit oder Begegnung) sowie die Gruppengröße verwendet. Als *Prädiktoren für die Beschreibung der konzeptionellen Zielsetzungen der Mitarbeitenden* werden analog zu den drei Kriterien das *Ziel Persönlichkeitsentwicklung* („Die TN sollen Gelegenheit haben, neue Seiten und Fähigkeiten bei sich zu entdecken“; „Die TN sollen neue Interessen und Hobbys entdecken können“; „Die Kreativität der TN soll gefördert werden“; $\alpha=.65$), das *Ziel politische Reflexion* („Die TN sollen Anstöße zu gesellschaftlichen/politischen Themen bekommen“) und das *Ziel interkulturelle Erfahrung* („Die TN sollen die Sprache des Gast-/Partnerlandes lernen und üben“; „Die TN sollen Alltag und Kultur der Gastregion kennen lernen, z.B. Essen, Musik, Traditionen“; „Die TN sollen in Kontakt mit Menschen aus der Gastregion kommen“; $\alpha=.78$) verwendet.

2.3 Das statistische Verfahren der Mehrebenenanalyse

Der hier verwendete Datensatz besteht aus Daten verschiedener Ebenen, die in einer geschachtelten Struktur vorliegen („nested design“). Prädiktoren für die untersuchten Kriterien werden sowohl auf individueller Ebene erwartet als auch auf Ebene der Gruppen. Der Einsatz der Mehrebenenanalyse erlaubt es, beide Ebenen gleich-

zeitig auf ihre Vorhersagekraft zu untersuchen (zum methodischen Hintergrund vgl. Kreft/De Leeuw 1998; Raudenbush/Bryk 2002; Ilg/Schweitzer 2010). Während ohne ein mehrebenenanalytisches Verfahren Effekte entweder über die individuellen Werte („complete pooling“) oder über die Gruppenmittelwerte („no pooling“) berechnet werden, erlaubt die Mehrebenenanalyse „partial pooling“, das beide Effekte kombiniert betrachtet (vgl. Gelman/Hill 2007: 252-254). Da keine spezifischen Annahmen im Blick auf variierende Steigungskoeffizienten zwischen den Gruppen vorliegen, verwendet der vorliegende Artikel nur Random Intercept-, aber keine Random Slope-Modelle.

Die hier dargestellte Untersuchung fokussiert vor allem auf den Zusammenhang zwischen pädagogischen Zielen und korrespondierenden Erfahrungen aufseiten der Jugendlichen. Die über individuelle Zielebewertungen erfassten Intentionen der Mitarbeitenden stellen den Indikator für die pädagogische Absicht des Teams dar. Daher gehen die jeweiligen Mittelwerte der Teams als Prädiktor in die Mehrebenenanalyse ein. Wie Tabelle 1 zeigt, stimmen die Ziele-Mittelwerte von durchschnittlich vier Mitarbeitenden pro Gruppe.

3. Ergebnisse der Mehrebenenanalyse

3.1 Unterschiedlichkeit der Gruppen und relevante Prädiktoren

Die erste Frage lautete: *Inwiefern unterscheidet sich das Erleben der Jugendlichen im Vergleich zwischen den einzelnen Gruppen?*

Einer Mehrebenenanalyse geht stets die als „Nullmodell“ bezeichnete Berechnung der Intraklassenkorrelation voraus. Sie verdeutlicht, wie groß der Varianzanteil ist, der zwischen den Gruppen liegt. Die Intraklassenkorrelation (ICC) in Tabelle 2 zeigt, dass diese Anteile mit einem Fünftel der Gesamtvarianz (Persönlichkeitsentwicklung und politische Reflexion) bzw. mehr als zwei Fünftel der Gesamtvarianz (interkulturelle Erfahrung) erheblich sind. Anders gesagt: Da die Effekte in den einzelnen Gruppen sich deutlich unterscheiden, ist eine mehrebenenanalytische Vorgehensweise angezeigt.

Tabelle 2: Intraklassenkorrelation (ICC) und Korrelationen für die Kriteriumsvariablen

	ICC	Korrelationen		
		1.	2.	3.
1. Persönlichkeitsentwicklung	.20		.27***	.36***
2. Politische Reflexion	.20	.34***		.34***
3. Interkulturelle Erfahrung	.43	.44***	.50***	

Anmerkungen: N=5.136 Teilnehmer(innen) von 242 Jugendgruppenfahrten (für interkulturelle Erfahrung: N=3.695 Teilnehmer(innen) von 229 Jugendgruppenfahrten). * $p < .05$, ** $p < .01$, *** $p < 0.001$. Oberhalb der Diagonalen sind die Korrelationen auf Individualebene, unterhalb der Diagonalen die Korrelationen der aggregierten Werte auf Gruppenebene abgedruckt.

Die zweite Frage lautete: *Welche Prädiktoren auf individueller Ebene und Gruppenebene zeigen Effekte auf die Erfahrungen der Jugendlichen bei den Fahrten?*

Die Bedeutsamkeit der Prädiktoren kann Tabelle 3 entnommen werden. Betrachtet wird hier zunächst das Modell 1, bei dem die pädagogischen Ziele noch nicht mit eingehen.

- Auf individueller Ebene ergibt sich ein statistisch signifikanter Zusammenhang des Alters mit Erfahrungen politischer Reflexion und mit interkulturellen Erfahrungen.
- Hinsichtlich des Geschlechts ist nur der Effekt statistisch signifikant, dass Jungen stärker von der Auseinandersetzung mit politischen Themen berichten als Mädchen.
- Die Vorerfahrung, also die Frage, ob die Jugendlichen bereits eine Freizeit oder internationale Jugendbegegnung besucht haben, wirkt sich nicht messbar auf die Kriteriumsvariablen aus. Da allerdings Vorerfahrungen in anderen Kontexten, z.B. Schullandheimen, nicht mit erfragt wurden, kann es sich hier auch um einen Effekt der zu unpräzisen Fragestellung handeln.
- Das Herkunftsland der Jugendlichen ist für die Kriterien unterschiedlich bedeutsam. Im Blick auf die Persönlichkeitsentwicklung profitieren Jugendliche, die aus den französischen, polnischen oder sonstigen Partnerländern kommen, deutlich mehr als deutsche Jugendliche, während die politische Reflexion offensichtlich eher typisch für Jugendliche aus Deutschland ist. Bei den interkulturellen Erfahrungen zeigen nur die Jugendlichen aus den ‚sonstigen‘ Ländern signifikant höhere Werte.
- Die Gruppengröße erweist sich als wenig relevanter Prädiktor, lediglich bei der politischen Reflexion zeigt sich ein negativer Zusammenhang zwischen der Zahl der Teilnehmenden in einer Gruppe und der politischen Reflexion während der Fahrt – und dieser tritt nur in Modell 1 auf.
- Erhebliche Unterschiede ergeben sich jedoch in Abhängigkeit vom jeweiligen Typ der Gruppenfahrt: Hinsichtlich aller drei untersuchten Kriterien liegen die Werte bei internationalen Jugendbegegnungen signifikant höher als bei Jugendfreizeiten. Dies ist insbesondere für interkulturelle Erfahrungen der Fall, bei denen der Vorsprung der Begegnungen vor den Freizeiten sogar eine ganze Standardabweichung beträgt.

Tabelle 3: Befunde aus Mehrebenenanalysen: Vorhersage der Persönlichkeitsentwicklung, der politischen Reflexion und der interkulturellen Erfahrung

	Persönlichkeitsentwicklung				Politische Reflexion				Interkulturelle Erfahrung			
	Modell 1		Modell 2		Modell 1		Modell 2		Modell 1		Modell 2	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
(Intercept)	-.24***	.05	-.22***	.06	-.14**	.05	.00	.06	-.59***	.05	-.41***	.07
<i>Individualvariablen</i>												
Alter	.01	.02	.02	.02	.12***	.02	.12***	.02	.04*	.02	.04*	.02
Differenz für Mädchen	.01	.03	.01	.03	-.10***	.03	-.10***	.03	-.01	.03	-.01	.03
Differenz für Jugendliche mit Vorerfahrung	-.03	.03	-.03	.03	-.02	.03	-.02	.03	.00	.03	.00	.03
Differenz für französische Jugendliche	.31***	.05	.29***	.05	-.08	.05	-.06	.05	.00	.07	-.02	.07
Differenz für polnische Jugendliche	.38***	.05	.37***	.05	-.37***	.05	-.36***	.05	.10	.07	.10	.07
Differenz für Jugendliche anderer Nationalität	.40***	.04	.41***	.04	-.13**	.04	-.14**	.04	.23***	.05	.23***	.05
<i>Allgemeine Gruppenvariablen</i>												
Differenz für internationale Begegnungen	.21***	.06	.16	.09	.46***	.06	.22*	.09	1.01***	.06	.72***	.09
Gruppengröße	-.05	.03	-.05	.03	-.11**	.03	-.06	.03	-.05	.04	-.03	.04
<i>Ziele der Mitarbeitenden (Gruppenebene)</i>												
Ziel Persönlichkeitsentwicklung			.19***	.03			-.02	.02			.02	.03
Ziel Politische Reflexion			-.01	.03			.23***	.03			.02	.03
Ziel Interkulturelle Erfahrung			-.01	.04			-.01	.04			.16***	.04
<i>Varianzaufklärung</i>												
R ² Individualebene	.02		.02		.01		.01		.00		.00	
R ² Gruppenebene	.29		.46		.43		.58		.75		.77	

Anmerkungen: N=5.136 Teilnehmende von 242 Jugendgruppenfahrten (interkulturelle Erfahrung: N=3.695 Teilnehmende von 229 Jugendgruppenfahrten). B=Regressionskoeffizient, SE=Standardfehler. *p<0,5, **p<0,01, ***p<0,001. Jugendliche ‚anderer Nationalität‘ sind alle Jugendliche, die weder aus Deutschland noch aus Frankreich noch aus Polen kommen.

3.2 Zusammenhang von pädagogischen Konzeptionen und den Erfahrungen der Jugendlichen

Für die nächsten beiden Fragen werden die Ergebnisse des in Tabelle 3 berichteten Modells 2 relevant: Frage 3: *Welcher zusätzliche Beitrag zur Varianzaufklärung ergibt sich, wenn die Ziele der Mitarbeitenden als konzeptionelle Prädiktoren in die Analyse mit eingehen?* Frage 4: *Zeigt sich für die untersuchten Kriteriumsbereiche ein spezifischer Einfluss der Mitarbeitendenziele, so dass konzeptionelle Schwerpunktlegerungen in einem bestimmten Bereich sich in den Rückmeldungen der Jugendlichen widerspiegeln?*

Bei allen drei Kriterien steigt die aufgeklärte Varianz von Modell 1 zu Modell 2 an. Der Unterschied in der Varianzaufklärung ist für die Kriterien Persönlichkeitsentwicklung und politische Reflexion erheblich, für die interkulturelle Erfahrung allerdings nur gering.

Mit der Integration der pädagogischen Ziele (Modell 2) schwächt sich die Differenz in Abhängigkeit vom Gruppenfahrt-Typ deutlich ab – ein Effekt, der daher stammt, dass die Ziele der Mitarbeitenden von Begegnungen sich von den Zielen der Freizeitmitarbeitenden unterscheiden und daher die unterschiedlichen Ziele einen Teil der Differenz in den Kriterien zwischen beiden Fahrt-Typen erklären.

Bei allen drei Kriterien erweist sich der Einfluss der korrespondierenden Mitarbeitendenziele als spezifisch: Das Ziel Persönlichkeitsentwicklung geht mit erhöhten Werten genau dieses Kriteriums einher, aber nicht mit den beiden anderen Kriteriumsvariablen. Entsprechendes gilt für die Ziele politische Reflexion und interkulturelle Erfahrung. Die Ziele der Mitarbeitenden zeigen demnach nicht einen diffusen Zusammenhang mit den ‚outcomes‘ einer Jugendgruppenfahrt, sondern tragen zur Vorhersage der entsprechenden Bildungswirkungen spezifisch bei.

4. Erfahrungen mit vernetzter Selbstevaluation

Die bislang beantworteten Fragen bezogen sich auf inhaltliche Aspekte, die mithilfe der nun zur Verfügung stehenden Datensätze beantwortet werden konnten. In einem nächsten Schritt soll reflektiert werden, welche Zwischenbilanz vier Jahre nach der Implementierung der vernetzten Selbstevaluation im Hinblick auf das Evaluationsverfahren gezogen werden kann.

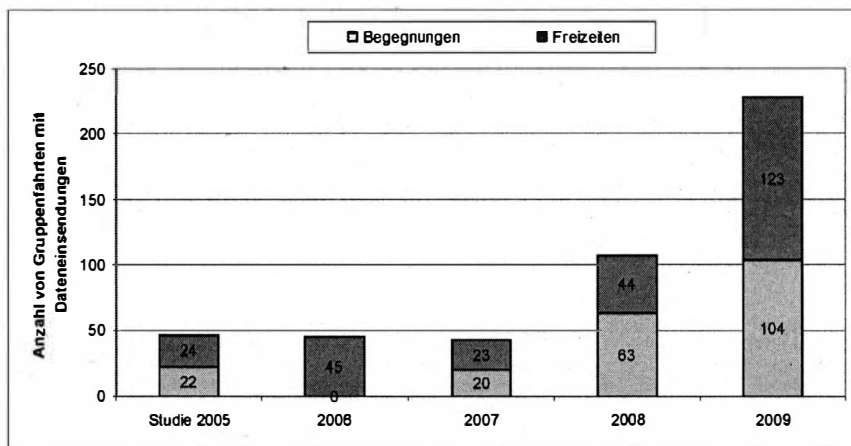
4.1 Datenrücklauf

Die fünfte Forschungsfrage lautete: *Welche Chancen und Grenzen ergeben sich für ein Konzept der vernetzten Selbstevaluation? Lässt sich bei einem lokal autonom einsetzbaren Evaluationssystem der Datenrücklauf zu einer zentralen Datensammelstelle gewährleisten? Wie kann die Motivation für die Dateneinsendung erhöht werden?*

Das Evaluationsverfahren, wie es für Freizeiten und internationale Jugendbegegnungen entwickelt wurde, belässt die Durchführung der Evaluation vollständig in der Autonomie der lokalen Jugendarbeits-Träger. Grund für diesen Autonomie-Ansatz war nicht nur die Vorgabe, ein kostenneutrales, also ohne externe Fachkräfte durchführbares Verfahren zu entwickeln, sondern auch die aus vielen Rückmeldungen von Praktiker(inne)n deutlich gewordene Sorge, dass ein zentral gesteuertes Evaluationsverfahren als Kontrollinstrument missbraucht und mit Sanktionen im Blick auf finanzielle Zuschüsse belegt werden könnte – oder zumindest mit entsprechenden Befürchtungen verbunden wäre. Bei der Entwicklung der Evaluationsinstrumente wurden daher in Anlehnung an die gängigen Evaluationsstandards (vgl. DeGEval 2008) Grundsätze für die Handhabung formuliert, die den Veranstaltern als Absicherung dafür dienen sollten, dass auch bei zentraler Dateneinsendung ihre Daten nicht ‚gegen sie verwendet‘ werden konnten. Die Grundsätze finden sich unter <http://www.freizeitenevaluation.de/grundsaeetze>, dort heißt es unter anderem: „Das Evaluationsinstrument wird ausschließlich zur Qualitätsentwicklung und nicht als ‚Überwachungsinstrument‘ verwendet“. Die gesammelten Daten werden daher von den Trägern per E-Mail an eine neutrale Stelle gesandt und von dort aus jeweils anonymisiert weitergeleitet. Eine vollständig anonyme Dateneinsendung über eine Internetseite wurde zwar eingerichtet, fand aber kaum Resonanz, da Träger, die auf die Angabe eines Absenders ganz verzichteten, nicht von der angebotenen Kompaktauswertung profitieren können. Die in Abbildung 2 dargestellten Rückläufe zeigen, dass die Dateneinsendung über die Jahre hinweg immer stärker voranging. Nach den Grundlagenstudien in 2005 konnte die Evaluation für Freizeiten bereits 2006 im Standardverfahren eingesetzt werden, die Evaluation für internationale Jugendbegegnungen erst ab 2007. Da etliche Träger die Evaluation durchführen ohne die Daten einzusenden, handelt es sich bei der Anzahl eingesandter Daten nur um die Spitze des Eisbergs – eines Eisbergs, über dessen wahre Größe allerdings leider keine Aussage getroffen werden kann.

Der zunächst erfreuliche Anstieg des jährlich eingesandten Datenvolumens darf allerdings nicht über Probleme hinwegtäuschen, die sich seit der Implementierung der vernetzten Selbstevaluation gezeigt haben. So besteht bei den Veranstaltern vor Ort vor allem die Motivation, diejenigen Fragebögen einzusetzen, von deren Auswertung man für die eigene Arbeit am meisten profitiert. Dies gilt für die Teilnehmer-Fragebögen sehr viel mehr als für die Mitarbeiter-Fragebögen. Nur für etwa die Hälfte der insgesamt 468 Maßnahmen liegen beide Fragebogen-Typen vor (vgl. Abschnitt 2.1). Die zusätzliche Erfassung der Rahmenbedingungen in einem Kurz-Fragebogen dauert nur wenige Minuten. Trotzdem wurden Daten häufig ohne diesen Erfassungsbogen eingesandt, was für die Verwertbarkeit der Daten auf lokaler Ebene zwar keinen Unterschied macht, aber zum Fehlen von Daten auf Gruppenebene führt. Mittlerweile ist die Einsendung der Rahmenbedingungen eine Vorbedingung für die Erstellung von Kompaktauswertungen, wodurch die Motivation zur Rücksendung deutlich verbessert werden konnte.

Abbildung 2: Anzahl von Jugendgruppenfahrten, die ihre Daten einsandten



Anmerkungen: Die Zahl der ausgewerteten Teilnehmer-Fragebögen liegt insgesamt bei 10.727 (2005: 1.386; 2006: 682; 2007: 941; 2008: 2.807; 2009: 4.911). Die Zahl der ausgewerteten Mitarbeiter-Fragebögen liegt insgesamt bei 1.227 (2005: 213; 2006: 57; 2007: 115; 2008: 270; 2009: 572). Die im Jahr 2010 eingegangenen Daten konnten für die Berechnungen im vorliegenden Artikel nicht mehr verwendet werden, ihre Anzahl wird daher nicht im Schaubild dargestellt: In 2010 wurden Fragebögen von 4.877 Teilnehmenden und 725 Mitarbeitenden von 131 Begegnungen und 80 Freizeiten eingesandt, die aktuellen Ergebnisse werden in Ilg/Dubiski 2011 veröffentlicht.

Ein wichtiger Grund für den erfreulich hohen Datenbestand ist darin zu sehen, dass die bundeszentralen und internationalen Institutionen, die das Evaluationsverfahren fördern (Deutsch-Französisches Jugendwerk, Deutsch-Polnisches Jugendwerk sowie transfer e.V. und IJAB – Fachstelle für Internationale Jugendarbeit der Bundesrepublik Deutschland, jeweils finanziert durch das BMFSFJ), die Implementierung mit Service-Angeboten begleiteten (Dokumentation vgl. Peters/Giebel 2011). So konnten sich Träger vor Ort individuell beim Aufbau einer Evaluationsstrategie unterstützen lassen oder die Dateneingabe der Fragebögen kostenfrei an extern finanzierte Honorarkräfte abgeben. Inwieweit die Evaluationsverfahren unter dem in der Jugendarbeit allgegenwärtigen Zeit- und Personaldruck auch dann durchgeführt werden, wenn solche Begleitung nicht gewährleistet ist, kann nicht abgeschätzt werden. Deutlich ist, dass Evaluation in non-formalen Settings noch keine Selbstverständlichkeit darstellt und daher Motivation und technische Unterstützung von außen wichtige Begleitfaktoren für das Gelingen vernetzter Selbstevaluation darstellen. Hier stellt sich für die zentralen Institutionen die Frage, welche Finanzmittel sie kontinuierlich bereitstellen wollen, um Daten über die Arbeit vor Ort zu erhalten und den Trägern zugleich Selbstevaluation und die damit verbundene Qualitätssicherung zu ermöglichen.

4.2 Gütekriterien

Die letzte Forschungsfrage bezieht sich auf die Verwendbarkeit der durch eine vernetzte Selbstevaluation erhaltenen Daten: *Sind die Daten, die in einer Selbstevaluation durch Fachkräfte ohne Evaluationsausbildung erhoben werden, hinsichtlich der Gütekriterien für eine wissenschaftliche Auswertung verwendbar?*

Zunächst muss hierbei gefragt werden, inwieweit die eingesandten Daten *repräsentativ* für das Forschungsfeld der Freizeiten bzw. internationalen Jugendbegegnungen insgesamt sind. Die Antwort darauf fällt ernüchternd aus: Ein Anspruch auf Repräsentativität kann in den Feldern der Jugendfreizeiten und internationalen Jugendbegegnungen aufgrund der Unstrukturiertheit des Feldes grundsätzlich nicht eingelöst werden. Im Bereich der Jugendfreizeiten existiert kein Überblick, wie viele solcher Fahrten in Deutschland durchgeführt werden, da prinzipiell jede Kirchengemeinde, jeder Verein oder Jugendclub eine Freizeit durchführen kann. Der für 2008 erstmals unternommene Versuch, einen Überblick über die Datenlage zu bieten, dokumentiert die Grenzen der Erfassbarkeit eindrucklich (vgl. Gleu/Kosmale 2009). Bei den Daten, die in der vernetzten Selbstevaluation zusammengetragen werden, kann nicht davon ausgegangen werden, dass sie die Vielzahl von Trägern repräsentativ abdecken. Vielmehr gibt es wenige große Träger, die viele Gruppenfahrten evaluieren, während insbesondere die kleinen Träger die Zeit für die Einarbeitung in das Evaluationsverfahren angesichts ihrer geringen jährlichen Zahl von Fahrten oftmals nicht leisten können oder wollen. Eine plausible Annahme ist, dass Träger die Daten ihrer Gruppenfahrten dann eher einsenden, wenn die Ergebnisse erfreulich sind, während negative Daten – auch wenn keine Sanktionen zu befürchten sind – eher zurückgehalten werden. Ein Screening zur Überprüfung dieser Hypothese (das allerdings nur Gültigkeit besitzt, wenn die Fahrten insgesamt über die Jahre hinweg als ähnlich gut beurteilt werden) kann über die Angaben der Jugendlichen zu ihrer Gesamtzufriedenheit vorgenommen werden: 29% der zwischen 2006 und 2009 eingesandten Daten haben bei der Angabe der Gesamtzufriedenheit durch die Jugendlichen einen Gruppenmittelwert von mehr als einem Zehntel Skalenpunkt negativer als der Mittelwert aus der entsprechenden Grundlagenstudie 2005, 15% bewegen sich auf ungefähr demselben Niveau wie dem der Grundlagenstudie, die restlichen 56% weisen einen um mindestens einen Zehntel Skalenpunkt positiveren Mittelwert auf. Dies kann zumindest als ein Hinweis darauf gedeutet werden, dass negative Ergebnisse weniger selbstverständlich an eine zentrale Stelle weitergegeben werden als positive Resultate.

Die Frage, ob sich die Gütekriterien der Instrumente beim Einsatz im nicht professionellen Kontext (also ohne geschultes Personal beim Verteilen der Fragebögen und bei der Dateneingabe) als stabil erweisen, kann über die *Cronbach-Alpha-Werte* überprüft werden. Verzerrungen durch schlechte Instruktionen, Tippfehler bei der Datenerfassung usw. müssten sich in geringeren Werten für die interne Konsistenz der Indizes niederschlagen. Die Ergebnisse in Tabelle 4 verdeutlichen, dass die Daten aus der vernetzten Selbstevaluation für die meisten Indizes über eine sehr ähnliche interne Konsistenz verfügen wie in der jeweiligen Grundlagenstudie aus dem Jahr 2005. Dies gilt fast durchweg für die Daten der Teilnehmenden, einge-

schränkt auch für die der Mitarbeitenden. Ein Grund dafür dürfte darin zu suchen sein, dass das Ausfüllen der Teilnehmer-Fragebögen sehr stark standardisiert ist (am vorletzten Tag der Fahrt, nach dem Frühstück), während für die Mitarbeiter-Befragung sowohl Zeitpunkte vor Beginn der Reise als auch in den ersten Tagen der Maßnahme möglich sind.

Tabelle 4: Cronbach-Alpha-Werte im Vergleich zwischen den Grundlagenstudien 2005 und der Phase der vernetzten Selbstevaluation 2006-2009

<i>Bezeichnung</i>	<i>Teilnehmer</i>		<i>Mitarbeiter</i>	
	<i>Grundlagen- studie</i>	<i>Selbst- evaluation</i>	<i>Grundlagen- studie</i>	<i>Selbst- evaluation</i>
<i>Freizeiten:</i>				
Sozialverhalten	.55	.53	.68	.65
Bezug zu den Mitarbeiter(inne)n	.68	.73	nur 1 Item	nur 1 Item
Kreativer Bereich	.53	.46	.58	.56
Ökologie	.65	.67	.81	.73
Partizipation	nur 1 Item	nur 1 Item	.76	.86
Allgemeine Bewertung	.78	.82	-	-
Zielklarheit der Mitarbeiter(innen)	-	-	.59	.69
<i>Internationale Jugendbegegnungen:</i>				
Soziales Lernen	.62	.65	.70	.54
Bezug zu den Mitarbeiter(inne)n	.78	.74	nur 1 Item	nur 1 Item
Kreativer Bereich	.42	.48	.45	.59
Fremdsprachenkompetenz	.57	.57	.70	.75
Kultur des Partnerlandes	.61	.59	.74	.73
Allgemeine Bewertung	.68	.79	-	-
Zielklarheit der Mitarbeiter(innen)	-	-	.76	.62

Anmerkungen: Die zugehörigen Items werden hier aus Platzgründen nicht abgedruckt. Sie sind in Ilg 2008a sowie Dubiski/Ilg 2008 zu finden. Einige Indizes ähneln sich zwischen Freizeiten und internationalen Jugendbegegnungen, sind aber zumeist nicht genau identisch.

Für die Verwertbarkeit der im Verfahren der vernetzten Selbstevaluation eingegangenen Daten muss insgesamt ein differenziertes Resümee gezogen werden: Die interne Konsistenz spricht insbesondere bei den Teilnehmer-Daten dafür, dass der Instrumenten-Einsatz auch bei einer Handhabung ohne ausgebildete Evaluator(innen) vor Ort nicht gefährdet ist. Es gibt allerdings deutliche Anzeichen dafür, dass Datensätze schwerpunktmäßig von solchen Gruppenfahrten eingesandt werden, die positive Ergebnisse erzielen. Ein Repräsentativitätsanspruch für die eingesandten Daten kann also zum jetzigen Stand nicht erhoben werden. Es erscheint insofern auch nicht als sinnvoll, die über Jahre eingesandten Daten als Vergleichswerte für die zukünftige Selbstevaluation zu veröffentlichen. Sollen die Vergleichsdaten aus 2005 aktualisiert werden, könnte dies jedoch über eine gezielte Auswahl von Trägern für die vernetzte Selbstevaluation geschehen, so dass die einbezogenen Daten ein möglichst repräsentatives Bild für die Jugendgruppenfahrten ergeben.

Auch ohne Repräsentativitätsanspruch bleibt die Auswertung der Daten aus der vernetzten Selbstevaluation für die wissenschaftliche Begleitung des Feldes interessant. Beispielsweise geben die soziodemografischen Daten zu den Teilnehmenden einen Anhaltspunkt dafür, ob die Teilnehmerschaft sich über die Jahre ändert. Auch gesellschaftliche Veränderungen schlagen sich in den Daten nieder: So liegt beispielsweise der Anteil von deutschen Jugendlichen, die bei einer Freizeit nach eigener Auskunft „mehr geraucht haben als zuhause“ in den Jahren 2005 und 2006 bei 19% der Befragten. Mit der im Jahr 2007 in Kraft getretenen gesetzlichen Verschiebung der Altersgrenze für Rauchen in der Öffentlichkeit von 16 auf 18 Jahre geht dieser Anteil im Jahr 2007 schlagartig zurück. Er liegt für das Jahr 2007 bei 7%, für 2008 bei 9% und steigt in 2009 schließlich auf 12% an.

5. Diskussion

5.1 Zusammenfassung der Ergebnisse

Die Ergebnisse der Mehrebenenanalyse zeigen, dass relevante Prädiktoren für die Vorhersage der Kriteriumsvariablen sowohl auf der Individual- als auch auf der Gruppenebene liegen. Während die Varianz auf Gruppenebene durch die Prädiktoren in hohem Maße aufgeklärt werden kann (insbesondere für die interkulturelle Erfahrung), bleibt die Varianzaufklärung auf individueller Ebene nur gering. Hier muss weiter nach individuellen Voraussetzungen gesucht werden, die in einem präaktiven Zusammenhang mit den Bildungserfahrungen der Jugendlichen stehen. Von den hier identifizierten Individualvariablen sind insbesondere die Länderzugehörigkeit, nicht aber die Vorerfahrung mit Gruppenfahrten und auch nur zu einem überraschend geringen Anteil Alter und Geschlecht von Relevanz.

Bei den Gruppenvariablen kommt dem Typ der Gruppenfahrt eine sehr bedeutende Rolle in der Vorhersage der Kriteriumswerte zu. Freizeiten und internationale Jugendbegegnungen sind demnach – trotz struktureller Ähnlichkeiten – in ihren Auswirkungen klar unterscheidbare Formen der kurzzeitpädagogischen Jugendarbeit. Ein deutlicher Unterschied findet sich vor allem bei der interkulturellen Erfahrung, aber auch hinsichtlich der politischen Reflexion. Dieses Ergebnis kann als eine Bestätigung der programmatischen Behauptung angesehen werden, dass internationale Jugendbegegnungen durch ihre direkte Auseinandersetzung mit verschiedenen Ländern und Kulturen sehr viel deutlicher zu Aspekten des Verständnisses für Fremdes beitragen als dies bei einer Gruppenreise von Jugendlichen aus demselben Land der Fall ist, selbst wenn diese Freizeit – wie es zumeist der Fall ist – ins Ausland führt.

Die als Indikatoren für das pädagogische Konzept verwendeten Ziele der Mitarbeitenden erweisen sich als hilfreiche Prädiktoren für die Varianzaufklärung auf Gruppenebene. Alle drei Zielsetzungen der Mitarbeitenden gehen mit entsprechend erhöhten Erlebnis-Rückmeldungen der Teilnehmenden einher. Es handelt sich bei allen drei Kriterien also um Erlebnisbereiche, bei denen das pädagogische Konzept, das sich in den vorab formulierten Zielen der Mitarbeitenden niederschlägt, in ei-

nem nachweisbaren Zusammenhang mit dem von den Jugendlichen am Ende berichteten Erleben der Gruppenfahrt steht. Dies stützt die Hypothese, dass Zielsetzungen, die sich auf einen individuellen Sachverhalt beziehen, stärker durch das pädagogische Personal determinierbar sind als solche, die sich auf Prozesse innerhalb der Gruppe, beispielsweise das Entstehen von Gemeinschaftserfahrung, beziehen. Für die pädagogische Praxis dürften solche Befunde insbesondere im Blick auf die Schulung von Mitarbeitenden relevant sein: Der sich in den Zielen niederschlagende ‚innere Wertekompass‘ der Mitarbeitenden hat offensichtlich eine prägende Bedeutung für das Geschehen bei einer Gruppenfahrt von Jugendlichen. Die individuellen Zielsetzungen sollten daher in der Vorbereitung einer Maßnahme thematisiert und mit denen des Veranstalters abgeglichen werden. Insgesamt müsste die Frage der individuellen Wert- und Zielhaltungen der einzelnen pädagogischen Mitarbeitenden im gesamten non-formalen Bildungsbereich stärker in den Fokus rücken, da ihr offensichtlich eine für die pädagogischen Effekte bislang nicht ausreichend berücksichtigte Bedeutung zukommt.

Die methodische Analyse zum Verfahren der vernetzten Selbstevaluation zeigt hinsichtlich der internen Konsistenz der Indizes eine hohe Übereinstimmung zwischen den Werten aus der Grundlagenstudie und denen der später selbständig eingesandten Daten. Dies spricht dafür, dass der Einsatz der Instrumente durch lokale Träger nicht zu grundsätzlichen Problemen bei der Datenerhebung führt. Der Datenrücklauf wuchs seit Einrichtung der vernetzten Selbstevaluation kontinuierlich an, erforderte jedoch Unterstützungssysteme, um den Veranstaltern bei der Durchführung der Evaluation möglichst wenig Aufwand zu bereiten. Eine repräsentative Datenbasis kann mit den eingesandten Daten nicht erzielt werden, dennoch ergeben sich aus den kumulierten Daten hilfreiche Erkenntnisse für die wissenschaftliche Begleitung eines bislang empirisch wenig greifbaren Bildungsfeldes. Mit über 15.000 Teilnehmerdaten von 679 Maßnahmen (inklusive der Daten aus 2010) ergab sich – mit begrenzten Fördermitteln – durch die vernetzte Selbstevaluation die größte bislang in der wissenschaftlichen Literatur dokumentierte Sammlung zu Jugendgruppenreisen.

5.2 Perspektiven der Weiterentwicklung

Der modulare Aufbau der hier vorgestellten Evaluation von Freizeiten und internationalen Jugendbegegnungen bietet vielfältige Möglichkeiten der Weiterentwicklung. Der Grundfragebogen mit wissenschaftlich abgesicherten Items kann einerseits durch Veranstalter vor Ort maßnahmenspezifisch durch eigene Items erweitert werden. Daneben wurde ein ‚Item-Pool‘ aufgebaut, der weitere methodisch abgesicherte Items für bestimmte inhaltliche Ausrichtungen der Fahrten zur Verfügung stellt, unter anderem für die Themen Gesundheitserziehung, künstlerische Arbeit oder religiöse Inhalte. Im Jahr 2010 wurde – neben der oben erwähnten kontinuierlichen Dateneinsendung im Jugendbereich – eine Weiterentwicklung des Evaluationsystems für Freizeiten von Kindern von 8 bis 12 Jahren (vgl. Peters et al. 2011) sowie für internationale Begegnungen im berufsbezogenen Bereich durchgeführt.

Für vertiefte Analysen zu den pädagogischen Prozessen bei Jugendgruppenfahrten sind die hier eingesetzten Evaluationsinstrumente allerdings nicht ausreichend differenziert. Hier wären Untersuchungen mit einem pre-camp, post-camp und follow-up Design wünschenswert, wie sie von Thurber et al. (2007) in den USA umgesetzt wurden. Auch eine zusätzliche Validierung der Items durch externe Einschätzungen der erzielten Bildungswirkungen (beispielsweise durch Eltern nach der Rückkehr der Jugendlichen in die Familie) bleibt in methodischer Hinsicht ein noch nicht eingelöstes Desiderat. Bei den Mitarbeitenden wären Beobachtungsstudien wünschenswert, um den Zusammenhang zwischen schriftlich geäußerten Zielen und konkreter pädagogischer Praxis genauer in den Blick nehmen zu können. Auch Untersuchungen mit einem quasi-experimentellen Design gehören zu den Desiderata der Evaluation von Gruppenfahrten, stoßen in der Praxis jedoch auf zahlreiche praktische und ethische Grenzen. Zudem fehlen im Bereich der non-formalen Bildung bislang oftmals die benötigten Finanzierungsmittel für groß angelegte wissenschaftliche Studien.

5.3 Die Evaluation non-formaler Bildung

Am Beispiel der Evaluation von Jugendgruppenfahrten wurden exemplarisch die Chancen und Herausforderungen für die Evaluation non-formaler Bildungsbereiche deutlich. Während sich die empirische Bildungsforschung in den letzten Jahren mit großen Evaluationsstudien (insbesondere PISA) intensiv mit der formalen Bildung beschäftigt, bleibt die außerschulische Jugendbildung bislang empirisch weitgehend unberücksichtigt. Dies zeigt sich auch im Mangel an Theorien über die in diesem Feld auftretenden Wirkmechanismen. Eine wichtige Aufgabe liegt – trotz der zunächst ermutigenden Ergebnisse von Henderson et al. (2007) – darin, weitere potenzielle Kontextfaktoren zu identifizieren, die für den Erfolg oder Misserfolg eines non-formalen Bildungsangebots verantwortlich sind. Die hier vorgelegten Analysen konnten insbesondere Aspekte der Programmkomponenten bei Gruppenfahrten aufgrund der geringen Zahl vorliegender Variablen auf Gruppenebene nicht berücksichtigen.

Der Einsatz der Mehrebenenanalyse als Verfahren zur simultanen Berücksichtigung von Einflussvariablen auf Individual- und Gruppenebene erweist sich für Gruppenangebote als ein wichtiges Analyse-Instrument und sollte zukünftig auch im non-formalen Bildungsbereich weitaus häufiger praktiziert werden als bislang. Eine wichtige Voraussetzung dafür sind allerdings große Datensätze (üblicherweise mit mehreren tausend Befragten). Ein Ansatz, wie solche Datensätze im non-formalen Bildungsbereich zu gewinnen sein können, ist mit dem Verfahren der vernetzten Selbstevaluation erfolgreich eingeführt. Trotz der benannten Grenzen dieser Weise der Datengewinnung erweist sich ein solches Verfahren als eine der wenigen praktikablen Möglichkeiten, auch unter der Vorgabe knapper finanzieller Mittel, verlässliche Datensammlungen zu generieren. Das Grundprinzip der vernetzten Selbstevaluation könnte in den nächsten Jahren auch für andere Felder non-formaler Bildung zu einem Schlüssel für die praxisnahe Begleitforschung werden.

Danksagung

Für die Förderung des hier vorgestellten Forschungsprojekts danken die Autoren folgenden Institutionen: Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ), Deutsch-Französisches Jugendwerk (DFJW), Deutsch-Polnisches Jugendwerk (DPJW), IJAB – Fachstelle für internationale Jugendarbeit, transfer e.V., Trainingsseminare für Kinder- und Jugendreisen und internationale Begegnungen (TiB) sowie dem Forscher-Praktiker-Dialog internationale Jugendarbeit.

6. Literatur

- Arbeitskreis G5 (Hg.) (2009): Das Wissen zur Kinder- und Jugendarbeit. Die empirische Forschung 1998-2008. Ein kommentierter Überblick für die Praxis. Neuss: LJR NRW. Verfügbar unter <http://www.forschungsverbund.tu-dortmund.de/index.php?id=100> [24.08.2011].
- Arbeitskreis G5 (Hrsg.) (2010): Kapuzenpulli meets Nadelstreifen. Die Kinder- und Jugendarbeit im Fokus von Wissenschaft und Wirtschaft. Neuss: LJR NRW. Verfügbar unter: <http://www.forschungsverbund.tu-dortmund.de/index.php?id=100> [24.08.2011].
- Beywl, Wolfgang (2006): Demokratie braucht wirkungsorientierte Evaluation – Entwicklungspfade im Kontext der Kinder- und Jugendhilfe. In: Projekt eXe (Hg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: Deutsches Jugendinstitut (DJI), S. 25-46.
- BMFSFJ – Bundesministerium für Familie, Senioren, Frauen und Jugend (2005): Zwölfter Kinder- und Jugendbericht. Bericht über die Lebenssituation junger Menschen und die Leistungen der Kinder- und Jugendhilfe in Deutschland. Bonn: Deutscher Bundestag.
- Bundesjugendkuratorium (2001): Zukunftsfähigkeit sichern! Für ein neues Verhältnis von Bildung und Jugendhilfe. Stellungnahme des Bundesjugendkuratoriums im Dezember 2001. Verfügbar unter: http://www.bundesjugendkuratorium.de/pdf/1999-2002/bjk_2001_stellungnahme_zukunftsfahigkeit_sichern.pdf [05.06.2011].
- DeGEval – Gesellschaft für Evaluation e.V. (2008): Standards für Evaluation. Mainz: DeGEval (4. Aufl.).
- Dubiski, Judith (2010): Institutionelle Diskriminierung in der Internationalen Jugendarbeit? In: IJAB (Hg.): Forum Jugendarbeit International 2008-2010. Internationale Jugendarbeit und Chancengleichheit. Bonn: IJAB, S. 385-397.
- Dubiski, Judith/Ilg, Wolfgang (Hg.) (2008): Evaluation Internationaler Jugendbegegnungen. Ein Verfahren zur Auswertung von Begegnungen. Berlin/Paris/Warschau: Deutsch-Französisches Jugendwerk (DFJW) und Deutsch-Polnisches Jugendwerk (DPJW).
- Garst, Barry A./Bruce, F. A. (2003): Identifying 4-H Camping Outcomes Using a Standardized Evaluation Process Across Multiple 4-H Educational Centers. In: Journal of Extension 41 (3). Verfügbar unter: <http://www.joe.org/joe/2003june/rb2.shtml> [05.06.2011].
- Gelman, Andrew/Hill, Jennifer (2007): Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge: Cambridge University Press.
- Gleu, Ritva K./Kosmale, Jens D. (Hg.) (2009): Deutsche Kinder- und Jugendreisen 2008. Aktuelle Daten zu Struktur und Volumen. Berlin: Bundesforum Kinder- und Jugendreisen.
- Henderson, Karla A./Bialeschki, M. Deborah/Thurber, Christopher A./Whitaker, Leslie S./Marsh, Paul E. (2007): Components of camp experiences for positive youth development. In: Journal of Youth Development 1, S. 17-28.
- Henderson, Karla A./Marsh, Paul E./Bialeschki, M. Deborah/Scanlin, Margery M./Whitaker, Leslie S./Thurber, Christopher A./Burkhardt, Mark (2006b): Intentional youth development through camp experiences. In: Camping Magazine, 79 (5), S. 6-8.
- Henderson, Karla A./Thurber, Christopher A./Whitaker, Leslie S./Bialeschki, M. Deborah/Scanlin, Margery M. (2006a): Development and Application of a Camper Growth Index for Youth. In: Journal of Experiential Education 29, S. 1-17.
- Ilg, Wolfgang (2005): Freizeiten auswerten – Perspektiven gewinnen. Grundlagen, Ergebnisse und Anleitung zur Evaluation von Jugendreisen im Evangelischen Jugendwerk in Württemberg. Bremen: IFKA (2. Aufl.).
- Ilg, Wolfgang (2007): Von der Gruppe ins Leben. Wirkungen der Jugendarbeit für Persönlichkeitsbil-

- dung und Lebensgestaltung – eine Sekundäranalyse. In: Ilg, Wolfgang/Weingardt, Martin (Hg.): Übergänge in der Bildungsarbeit mit Jugendlichen. Empirische Studien zu den Nahtstellen von Jugendarbeit, Schule und Freizeit. Weinheim/München: Juventa, S. 171-200.
- Ilg, Wolfgang (2008a): Evaluation von Freizeiten und Jugendreisen. Einführung und Ergebnisse zum bundesweiten Standard-Verfahren. Hannover: aej.
- Ilg, Wolfgang (2008b): Freizeitevaluation. Hintergründe zu Methodik und Einsatz des Standard-Verfahrens für die Auswertung von Freizeiten und internationalen Jugendbegegnungen. In: deutsche jugend 56, S. 101-106.
- Ilg, Wolfgang (2010): Freizeitevaluation. Entwicklung eines Evaluationskonzepts für Freizeiten in der kirchlichen Jugendarbeit. Saarbrücken: VDM-Verlag.
- Ilg, Wolfgang/Dubiski, Judith (2011): Begegnung schafft Perspektiven. Empirische Einblicke in internationale Jugendbegegnungen. Berlin/Paris/Warschau: Deutsch-Französisches Jugendwerk (DFJW) und Deutsch-Polnisches Jugendwerk (DPJW).
- Ilg, Wolfgang/Lüdtke, Oliver (2011): Prädiktoren von Bildungsprozessen in der Konfirmandenarbeit. Eine mehrbenenanalytische Untersuchung non-formaler Jugendbildung. In: Zeitschrift für Erziehungswissenschaft 14, S. 309-328.
- Ilg, Wolfgang/Schweitzer, Friedrich (2010): Researching confirmation work in Europe: The need for multi-level analysis for identifying individual and group influences in non-formal education. In: Journal of Empirical Theology 23, S. 159-178.
- Kloosterman, Paul/Giebel, Kerstin/Senyuva, Ozgehan (2007): T-Kit Nr. 10. Educational Evaluation in Youth Work. Strasbourg: Council of Europe Publishing.
- Kreft, Ita/De Leeuw, Jan (1998): Introducing multilevel modeling. London: Sage.
- Lüders, Christian/Haubrich, Karin (2006): Wirkungsevaluation in der Kinder- und Jugendhilfe: Über hohe Erwartungen, fachliche Erfordernisse und konzeptionelle Antworten. In: Projekt eXe (Hg.): Wirkungsevaluation in der Kinder- und Jugendhilfe. Einblicke in die Evaluationspraxis. München: Deutsches Jugendinstitut (DJI), S. 5-23.
- MGFFI – Ministerium für Generationen, Familie, Frauen und Integration des Landes Nordrhein-Westfalen (Hg.) (2007): Wenn Jugendliche einander begegnen. Evaluation des multinationalen und integrativen Jugendbegegnungsprojektes EIGHT COLUMNS. Düsseldorf: MGFFI.
- Müller, Burkhard (1996): Evaluation internationaler Begegnungen. Eine erste perspektivische Beleuchtung der Probleme von Auswertungsmethoden im Bereich des interkulturellen Lernens. Bad Honnef: Deutsch-Französisches Jugendwerk (DFJW).
- Müller, Werner (Hg.) (o.J.): Praxishandbuch Kinder- und Jugendfreizeiten. Loseblattwerk. Landsberg: mvg.
- Peters, Heike/Giebel, Kerstin (2011): Wege in eine Evaluationskultur. Evaluations-Coaching als Instrument zur Qualitätsentwicklung für Anbieter von Kinder- und Jugendfreizeiten sowie Jugendbegegnungen im In- und Ausland. Handreichung zu den Verfahren „Evaluation von Freizeiten“ und „Evaluation Internationaler Jugendbegegnungen“. Bonn/Köln: IJAB/transfer e.V.
- Peters, Heike/Otto, Stephanie/Ilg, Wolfgang/Kistner, Günter (2011, im Druck): Evaluation von Kinderfreizeiten. Wissenschaftliche Grundlagen, Ergebnisse und Anleitung zur eigenen Durchführung. Hannover: aej.
- Raudenbush, Stephen W./Bryk, Anthony S. (2002): Hierarchical Linear Models: applications and data analysis methods. Thousand Oaks/CA: Sage (2. Aufl.).
- Rauschenbach, Thomas (2009): Zukunftschance Bildung. Familie, Jugendhilfe und Schule in neuer Allianz. Weinheim/München: Juventa.
- Rauschenbach, Thomas/Bormann, Stefan/Düx, Wiebken/Liebig, Reinhard/Pothmann, Jens/Züchner, Ivo (2010): Lage und Zukunft der Kinder- und Jugendarbeit in Baden-Württemberg. Eine Expertise. München: Deutsches Jugendinstitut (DJI).
- Sturzenhecker, Benedikt/von Spiegel, Hiltrud (2009): Was hindert und fördert Selbstevaluation und Wirkungsreflexion in der Kinder- und Jugendarbeit? In: Lindner, Werner (Hg.): Kinder- und Jugendarbeit wirkt. Aktuelle und ausgewählte Evaluationsergebnisse der Kinder- und Jugendarbeit. Wiesbaden: VS Verlag (2. Aufl.), S. 309-321.
- Thomas, Alexander/Chang, Celine/Abt, Heike (2007): Erlebnisse, die verändern. Langzeitwirkungen der Teilnahme an internationalen Jugendbegegnungen. Göttingen: Vandenhoeck & Ruprecht.
- Thurber, Christopher A./Scanlin, Margery M./Whitaker, Leslie S./Henderson, Karla A. (2007): Youth development outcomes of the camp experience: Evidence for multidimensional growth. In: Journal of youth and adolescence; a multidisciplinary research publication 36, S. 241-254.
- Zitzmann, Christina (2007): „Das habe ich im Gefühl“. Perspektiven für Praxis und Wissenschaft der Wirkungsforschung. In: deutsche jugend 55, S. 170-177.