

Dr. Harald Tornow
Berlin, 21.09.2013

**ABSCHLUSSBERICHT FÜR DAS PROJEKT
UNTERSUCHUNG ZUR BESTIMMUNG STEUERUNGSRELEVANTER WIRKUNGS-
FAKTOREN IM HILFEPLANPROZESS IM RAHMEN DES GESAMTSTÄDTISCHEN
FACHCONTROLLINGS HILFEN ZUR ERZIEHUNG IN BERLIN**

Inhalt

1	Projekt	5
1.1	Zielsetzung	5
1.2	Projektbeschreibung.....	5
1.3	Projektablauf	6
2	Theoretisch-methodische Grundlagen für eine evidenzbasierte Steuerung von Hilfen zur Erziehung.....	8
2.1	Die Bedeutung der Steuerung im Qualitätsmanagement sozialer Dienstleistungen.....	8
2.2	Steuerungsebenen.....	9
2.3	Kennzahlen als Instrument für Steuerung.....	11
3	Messung von Wirkung und Wirksamkeit - am Beispiel der WIMES-Methode.....	13
3.1	Unterschiedliche Evaluationsverfahren	13
3.2	Wirkungsmessung mit der WIMES-Methode	15
4	Empirische Ergebnisse der Wirkungsevaluation (aggregiert über alle 5 Projektbezirke)	19
4.1	Datenbasis.....	19
4.2	Statistische Auswertungen (Lesehilfe)	20
4.2.1	Deskriptive Statistik.....	20
4.2.2	Explorative Statistik	21
4.3	Allgemeine Effektstärke aller Erziehungshilfen	27
4.4	Retrospektive Einschätzung der Zielerreichung der Hilfe	30
4.5	Retrospektive Beurteilung der Entwicklung während der Hilfe.....	30
4.6	Der Anteil der nicht planmäßig beendeten Hilfen (Abbruchquote)	30
4.7	Folgehilfen	31
4.8	Weitere mögliche Wirkungskennzahlen, die im Pilotprojekt nicht erhoben worden sind.....	33
4.9	Leistungstypen und Hilfearten differenzierte Wirkungskennzahlen	35
4.9.1	Alle ambulanten therapeutischen Hilfen.....	36
4.9.2	Im Einzelnen: ambulante Psychotherapie (Leistungstyp 1)	37
4.9.3	Im Einzelnen: integrative Lerntherapie (Leistungstyp 2)	38
4.9.4	Im Einzelnen: ambulante Familientherapie (Leistungstyp 3)	39
4.9.5	Ambulante sozialpädagogische Hilfen	40
4.9.6	Im Einzelnen: Soziale Gruppenarbeit (Leistungstyp 4).....	41
4.9.7	Im Einzelnen: Betreuungshilfe / Erziehungsbeistand (Leistungstyp 5)	42
4.9.8	Im Einzelnen: sozialpädagogische Familienhilfe (Leistungstyp 6)	43

4.9.9	Im Einzelnen: ambulante ISE (Leistungstyp 7)	44
4.9.10	Teilstationäre Hilfen in Tagesgruppen (Leistungstyp 8)	45
4.9.11	Stationäre Hilfen	52
4.9.12	Im Einzelnen: stationär (Gruppenangebot) (Leistungstyp 9).....	59
4.9.13	Im Einzelnen: stationär (Gruppenangebot Wohngemeinschaft) (Leistungstyp 10)	61
4.9.14	Im Einzelnen: stationär (familienanalog) (Leistungstyp 11)	62
4.9.15	Im Einzelnen: stationär (Individualangebot / BEW) (Leistungstyp 12)	64
4.9.16	Im Einzelnen: stationäre ISE (Leistungstyp 13).....	65
4.10	Wirkungsevaluation bei Clearingfällen (KCD = Krisenmanagement, Clearing, Diagnose)	66
4.11	Kooperative Wirkungsevaluation von Leistungsträger und Leistungserbringer	68
5	Implementierung wirkungsorientierter Kennzahlen - Praxiserfahrungen aus Sicht des e/l/s-Instituts	72
6	Empfehlungen zu einer kontinuierlichen Wirkungsevaluation und einer wirkungsorientierten Steuerung von Hilfen zur Erziehung in Berlin	75
7	Anlage Erhebungsinstrument	81
8	Anlage Datengrundlage	86
9	Anlage Liste der Auswertungen	87
10	Anlage Abbildungen und Tabellen (im Text)	91
11	Anlage Literaturverzeichnis.....	93
12	Anlage Glossar	95

Vorbemerkung

Der vorliegende Bericht fasst die wesentlichen Ergebnisse des Projektes zusammen. Die Zielsetzung und der Ablauf des Projektes sind im ersten Kapitel dargestellt.

Die folgenden beiden Kapitel befassen sich mit Theorie und Methodik von Wirkungsevaluation, evidenzbasierter Qualitätsentwicklung und wirkungsorientierter Fallsteuerung. Das eingesetzte Verfahren WIMES („Wirksamkeit messen“) wird ausführlicher besprochen, damit nachvollzogen werden kann, auf welche Weise die empirischen Ergebnisse zustande gekommen sind.

Den umfangreichsten Teil nimmt das Kapitel 4 ein, in dem die wesentlichen empirischen Befunde dargestellt sind und Hinweise gegeben werden, welche Ergebnisse steuerungsrelevant sind und entsprechend beachtet und weiter verfolgt werden sollten. Damit auch statistische „Laien“ dieses Kapitel verstehen können, wurde die Erklärung wichtiger statistischer Fachbegriffe vorangestellt. Hierzu verweisen wir auch auf den Anhang Seite 95. Die Ergebnisse sind pro Hilfeart / Leistungstyp tabellarisch wiedergegeben, um eine schnelle Orientierung zu erleichtern. Weitere Auswertungen liegen der Senatsverwaltung und den beteiligten Bezirken ebenso vor wie die erhobenen Rohdaten.

Die folgenden beiden Kapitel 5 und 6 sind aus subjektiver Sicht des mit dem Projekt beauftragten Institutes e/l/s (entwickeln - lernen - sichern) geschrieben. Wir kommen damit dem Auftrag nach, nicht nur die empirischen Datenergebnisse wiederzugeben, sondern Anregungen und Empfehlungen zu geben, welcher Nutzen und Erkenntnisgewinn aus den Projektergebnissen gezogen werden kann und wie die Implementierung einer verstärkten Wirkungsorientierung von Hilfen zur Erziehung in Berlin aussehen könnte.

1 Projekt

1.1 Zielsetzung

Das Abgeordnetenhaus hat die Senatsverwaltung für Bildung, Jugend und Wissenschaft mit Beschluss vom 11.06.2009 beauftragt, ein gesamtstädtisches Fachcontrolling für die Hilfen zur Erziehung (HzE) einzurichten (vgl. Drs. 16/2474). Ziele eines gesamtstädtisch standardisierten Fachcontrollings sind die Herstellung von Transparenz über die Fall- und Ausgabenentwicklung, die Analyse der Rahmenbedingungen und Zusammenhänge und die weitere Qualifizierung der Hilfeplanung.

Gemeinsam mit den Bezirken werden in sieben Arbeitspaketen die steuerungsrelevanten inhaltlichen, strukturellen und finanziellen Aspekte eines integrierten Fach- und Finanzcontrollings im Bereich Hilfe zur Erziehung erarbeitet bzw. fortentwickelt. Die extern begleitete Untersuchung der Hilfeplanung und Hilfegewährung unter dem Aspekt des Ziel-Wirkungszusammenhangs in fünf Jugendämtern ist ein zentrales Arbeitspaket.

Durch die systematische Untersuchung der Fallverläufe anhand der zentralen Zieldimensionen sollen Wirkungshinweise im Hilfeplanprozess zugeordnet und weiter ausgewertet werden. Die Untersuchung zielt auf konkrete fachliche Verbesserungen in der alltäglichen Fallarbeit und auf die Qualifizierungsmöglichkeiten bei der Entscheidung und Zielformulierung in der Hilfeplanung aufgrund gesicherter Erkenntnis über die Wirkdimensionen von Hilfen.

Die vorgelegten Ergebnisse können das Fundament für eine vertiefte regionale bzw. sozialraumbezogene Analyse und Bewertung sowie für eine bedarfsgerechte Weiterentwicklung der Hilfen zur Erziehung in Berlin sein.

1.2 Projektbeschreibung

In dem Pilotprojekt (2010 bis 2013) der Berliner Senatsverwaltung wurden Methoden und Verfahren erprobt, mit denen Kennzahlen zur fachlichen Steuerung von Hilfen zur Erziehung erhoben werden. Wirkungsindikatoren sollen zum einen das Fach- und Finanzcontrolling in den Organisationen mit Fakten und Daten untermauern, um Bedarfen mit effektiven und effizienten Hilfen begegnen zu können. Zum anderen soll die konkrete Fallsteuerung im Einzelfall wirkungsorientiert ausgestaltet werden. Es geht also um den Nachweis von Effektivität oder ihrer Verbesserungsnotwendigkeit und um eine Optimierung der Effizienz sozialer Dienstleistungen mit einem hohen Nutzwert für die Klienten.

Zur wissenschaftlichen Fundierung eines an Wirkungsindikatoren ausgerichteten Fachcontrollings der Hilfen zur Erziehung wurden Daten für den Zeitraum vom 1.1.2011 bis zum 31.12.2012 in fünf Bezirken (Mitte, Friedrichshain-Kreuzberg, Pankow, Marzahn-Hellersdorf, Steglitz-Zehlendorf) Daten zur Wirksamkeit von Hilfen zur Erziehung nach §§ 27 ff SGB VIII dokumentiert. Ausgeschlossen wurden Hilfen nach § 28 (Erziehungsberatung) und nach § 33 (Familienpflege). Clearing- und Krisenmanagementleistungen wurden ebenfalls evaluiert. Sie werden aber gesondert ausgewertet, genauso wie Hilfen für Mutter-Kind nach § 19, die nicht Gegenstand dieses Berichtes sind.

Die Daten wurden von den fallverantwortlichen SozialarbeiterInnen in den RSD zu drei Zeitpunkten erhoben, und zwar bei allen Neufällen bei der Bedarfserhebung vor Hilfebeginn und bei der Überprüfung der Eingangsdagnostik nach ca. 6 Wochen. Alle in dem Zeitraum beendeten Hilfen wurden hinsichtlich der zu dem Zeitpunkt feststellbaren Ressourcen/Kompetenzen versus Problembelastungen/Risiken eingeschätzt. Dass die Klienten sowohl bei der Bedarfsfeststellung und Zielsetzung der Hilfe als auch bei der Abschlussbewertung beteiligt wurden, setzten wir als selbstverständlichen fachlichen Standard voraus. Eine gesonderte Nutzerbefragung wurde (noch) nicht durchgeführt.

Der vorliegende Bericht fasst die Ergebnisse des Projektes zusammen. Für die empirischen Analysen werden alle Daten aus den fünf Bezirken aggregiert. Details aus den Bezirken werden nicht dargestellt, allenfalls als Hinweise auf den Grad der Unterschiedlichkeit zwischen den Bezirken. Die Ergebnisse auf der Bezirks- und soweit es möglich war, auf der Regionalebene sind in Auswertungen, Berichten und Präsentationen den Leitungskräften und MitarbeiterInnen in den jeweiligen Bezirken zurückgemeldet worden. In den Bezirksberichten sind die Gesamtergebnisse der 5 Bezirke als Berlin interner Benchmark genutzt worden, so dass die Ak-

teure in den Bezirken ihre Ergebnisse in Relation zu Berlin (natürlich nur in dem Ausschnitt der 5 Bezirke) einordnen und bewerten können.

1.3 Projektablauf

Der erste Durchlauf für eine Wirkungsevaluation wurde auf den Zeitraum vom 1.7.2010 bis zum 31.12.2011 festgelegt. Da die Sommerferien in die Anlaufphase fielen, stand dem Projekt in der ersten Phase de facto nur der Zeitraum vom 1.9.2010 bis Ende 2011 zur Verfügung.

Üblicherweise braucht es eine Vorbereitungszeit von mindestens einem halben Jahr, bevor alle organisatorischen und technisch-materiellen Voraussetzungen geschaffen sind und bevor die Mitarbeiter, die später die Dokumentationsaufgaben haben, auf das Projekt eingestimmt sind und die nötigen Informationen und Hilfsmittel haben. Außerdem braucht es dann noch etwa 3-6 Monate, bis sich Routinen einspielen und die Dokumentation „rund läuft“. Danach braucht es ein Jahr Dokumentation, um eine ausreichend große und differenzierte Stichprobe zu bekommen für eine hinreichend zuverlässige Querschnittuntersuchung¹. Es muss beachtet werden, dass für einen relativ kurzen Zeitraum von weniger als einem Jahr eine Längsschnittuntersuchung nicht möglich ist. Für eine prospektive Längsschnittevaluation bedarf es einer ausreichend umfangreichen Stichprobe, bei der in jedem Einzelfall Evaluationsdaten vom Hilfebeginn und vom Hilfeende vorliegen. Wählt man einen zu kurzen Erhebungszeitraum, kommt es einem systematischen Fehler, weil nur Hilfen mit kürzeren Laufzeiten untersucht werden

Um auch länger laufende Hilfen für eine Längsschnittuntersuchung zu erfassen und weil die Stichprobenumfänge und die Datenqualität am Ende der ersten Projektphase noch nicht befriedigend waren, wurde eine zweite Projektphase vom 1.1.2012 bis 31.12.2013 beschlossen. In dieser zweiten Phase sollten weiterhin alle Neufälle und beendeten Fälle mit dem Evaluationsverfahren untersucht und dokumentiert werden. Darüber hinaus sollte mit einigen Leistungserbringern eine gemeinsame kooperative Wirkungsevaluation erprobt werden und die Implementierung wirkungsorientierter Steuerung in das Hilfeplanverfahren vorbereitet werden.

Die zweite Projektphase (2012/2013) wurde dadurch modifiziert, dass aufgrund von Datenschutzauflagen des Unabhängigen Berliner Datenschützers umfängliche technische und organisatorische Änderungen bei der Datenerfassung vorgenommen werden mussten, die zum 31.10.2012 erfolgreich abgeschlossen werden konnten.

Am 17.1.2013 beschlossen die Senatsverwaltung und die Projektbezirke aufgrund der entstandenen Zeitverzögerung und erheblicher personeller Engpässe in den bezirklichen Jugendämtern, die Datenerfassung zum 30.5.2013 mit der Konsolidierung der Daten zum Stichtag 31.12.2012 zu beenden und den Projektschwerpunkt auf vertiefte Auswertungen und Analysen zu verlagern

Die Verbesserung der Datenqualität und die Vergrößerung des Stichprobenumfangs für eine Querschnittuntersuchung wurden erreicht und bilden die Basis der Ergebnisse, die in dem vorliegenden Bericht dargestellt werden. Außerdem wurden die Schnittstellen zwischen Wirkungsevaluation und dem Berliner Hilfeplanverfahren, den Berliner Qualitätsentwicklungsdialogen und der Sozialraumorientierung diskutiert und abgeglichen. Nicht erreicht wurden die Basis für eine Längsschnittanalyse, die den Datenzeitraum von drei Jahren erfordert hätte und der systematische Einbezug von Leistungserbringern in einer Phase systematischer kooperativer Datenerfassung.

¹ Die Unterscheidung zwischen einer Querschnittuntersuchung und einer Längsschnittuntersuchung wird im Methodenteil ausführlich erklärt. Der wesentliche Unterschied liegt darin, dass bei einem Querschnittdesign die Daten zweier unterschiedlicher Gruppen verglichen werden, nämlich diejenigen, die in einem bestimmten Zeitraum mit einer Hilfe starten, mit denjenigen, die in dem Zeitraum die Hilfe beenden. Bei einer Längsschnittuntersuchung werden Hilfen von Anfang bis zum Ende dokumentiert und in ihrem Verlauf bewertet. Die Dokumentation für Längsschnittuntersuchungen dauert bei Hilfen zur Erziehung mindestens drei Jahre.

Hier noch einmal ein Überblick über die Chronologie des Projektes:

Zeitraum	Teilziele und Vorgehen
Juli bis September 2010	Aufbau der Projektstruktur; Festlegung der teilnehmenden Bezirke; Festlegung der Koordinatoren in den Bezirken; technisch-methodische Vorbereitung; Vorbereitung der Informationsveranstaltungen und Schulungen in den Bezirken; Anpassung an die Berliner Dokumentationsstruktur (Leistungstypen, Sonderabfragen)
September bis Dezember 2010	Schulung der MitarbeiterInnen und Probelauf der Dokumentation (in Marzahn-Hellersdorf begann die Echt-Dokumentation bereits am 1.10.2010)
Januar bis November 2011	Echtdokumentation in den Bezirken; Einzelberatung und Nachschulung; erste Probeauswertung und deren Darstellung in den Bezirken; erste Zwischenabschätzung auf dem Fachtag am 14.6.2011; Information der Träger in den Bezirken Friedrichshain-Kreuzberg, Pankow und Marzahn-Hellersdorf; Schulung der Träger sowie Einrichten der Zugänge zum Dokumentationssystem
November 2011 bis Februar 2012	Zwischenabschätzung des bisherigen Datensatzes (Stichtag 10.11.2011); Präsentation der Ergebnisse und Workshops in den Bezirken in Gruppen unterschiedlicher Zusammensetzung; Aushändigen der Auswertungen als Dateien und in kommentierender Berichtsform; Information von Gremien auf bezirksübergreifender Ebene; Abschlussbericht der ersten Phase; Fachtag in Schloss Glienicke am 15.2.2012; Verlängerung des Projektes mit Erweiterung auf Hilfeplanung und Kooperation mit den Leistungserbringern
März 2012 bis Oktober 2012	Fortsetzung der Datenerhebung wie bisher und Vorbereitung auf die Implementierung in das Hilfeplanverfahren; Lösung der Datenschutzerfordernisse für das WIMES-Webportal, auch Datensicherheit nach BSI-Grundschutz; technische Vorbereitung der Daten-Verschlüsselung; Schulung der Leistungserbringer; Weiterentwicklung der WIMES-Dimensionen und des Mutter-Kind-Moduls
November 2012 bis April 2013	Wiederaufnahme der Dokumentation stockt aus technischen Gründen in einigen Bezirken, in anderen geht die Dokumentation weiter; Zwischenabschätzungen in einigen Bezirken; Stoppen der Dateneingabe, Nacharbeiten der Datenbestände 2012; Schließen des Datenbestandes zum 1.5.2013
Mai bis September 2013	Kommunikation zum Projekt mit Trägern und der Liga; Auswertungen für die Senatsverwaltung (alle 5 Bezirke) und die Bezirke im Einzelnen; Planung vertiefender Analysen; Abgabe des Abschlussberichtes mit Empfehlungen
September 2013 bis Dezember 2013	Mitarbeiterbefragung; Untersuchung stationärer Abbrüche; Diskussion der Bezirksergebnisse; abschließender Fachtag am 4.12.2013 in Schloss Glienicke

2 Theoretisch-methodische Grundlagen für eine evidenzbasierte Steuerung von Hilfen zur Erziehung

2.1 Die Bedeutung der Steuerung im Qualitätsmanagement sozialer Dienstleistungen

Der zentrale Begriff, um den in der sozialen Arbeit und hier speziell in der Jugendhilfe seit fast 20 Jahren geredet wird, ist der „Steuerung“. Gemeint sind damit Modelle und Verfahren, mit denen man auf unterschiedlichen Verantwortungsebenen gezielt Einfluss auf etwas nehmen kann, das offensichtlich nicht so recht zu planen und zu lenken ist.

Qualitätsmanagement ist die Summe der planenden, steuernden und prüfenden Aktivitäten, mit denen die Erfordernisse eines Produktes oder einer Dienstleistung möglichst effektiv und effizient erfüllt werden sollen (Bruhn, 2003) (Beckmann, 2004) (Dukek & Burmeister, 2012) (Eversheim, 1997) (Meinholt, 1996) (Ebel, 2001) (Walter Geiger & Willi Kotte, 2008) (Peterander & Speck, 1999). Es umfasst in allen Branchen, die Kundenerwartungen treffen wollen und müssen, Maßnahmen der Qualitätsentwicklung und der Qualitätssicherung. Qualitätsentwicklung meint, die Organisation so auszurichten, dass die Qualitätserfordernisse möglichst gut erfüllt werden, durch Prozesse der Produktplanung und der Produktion sowie durch einen kontinuierlichen Verbesserungsprozess, der insbesondere durch die Fakten und Daten zur Ergebnisqualität gespeist wird. Qualitätssicherung meint, die Ressourcen und Prozesse so zu lenken, dass ein Optimum an Ergebnisqualität erreicht wird. Nach den Modellen umfassender Qualitätsmanagementsysteme werden die Ergebnisse bei den Nutzern und die finanziellen Ergebnisse gemessen und bewertet und in die Organisation als Lernimpulse zurückgemeldet.

In der Untersuchung zur Erprobung von Wirkungskennzahlen in Berlin geht es genau um dieses Verständnis. Steuerung bedeutet in diesem Sinne nicht die retrospektive Erklärung von Wirkungen oder Abweichungen, sondern die Generierung eines Wissens zu proaktiver Einflussnahme auf ein System zur Erreichung von besten Ergebnissen zum kleinstmöglichen Preis (Breyer & Zweifel, 1992) (Bruhn, 1999) (Beckmann, 2004) (Jessen u. a., 2011) (Landes, 2011). Steuerung in diesem Sinne heißt, systematisch Erfahrungswissen dafür zu nutzen, in Zukunft Gelingenswahrscheinlichkeiten zu erhöhen oder Fehler- oder Misserfolgswahrscheinlichkeiten zu minimieren.

Im Berliner Max-Planck-Institut für Bildungsforschung beschäftigen sich derzeit Wissenschaftler mit der Steuerung komplexer Systeme unter den Bedingungen von Unbestimmtheit (Gigerenzer, 2011). Letztlich lassen sich Gewissheiten nicht generieren, wohl aber lassen sich Wahrscheinlichkeiten zukünftiger Ereignisse berechnen und damit Grundlagen für Entscheidungen in ungewissen Umwelten finden. Die Steuerungsfähigkeit hängt bei komplexen Systemen unter den Bedingungen der Ungewissheit stark davon ab, wie gut die Steuerenden in der Lage sind, mit Wahrscheinlichkeiten umzugehen („risk literacy“) (Gigerenzer, 2008). Interessanterweise ist es nicht der Umfang der Informationen, der Entscheidungen besser macht. Dieses widerspricht einem Trend in der Sozialen Arbeit, dass immer mehr Daten und immer komplexere Modelle gesucht werden, um zu angeblich besseren Ergebnissen zu führen. Die Empfehlung geht dahin, die entscheidenden Informationen zu nutzen und daraus praktikable, einfache und effiziente Faustregeln abzuleiten (Heuristiken). Das einzige Kriterium, das erfüllt werden muss, ist das der Nützlichkeit.

Steuerung setzt eine klare Definition von Zielen voraus. In dem Berliner Projekt gingen wir davon aus, dass die Ziele von Erziehungshilfe im Grundsatz beschrieben sind. Sie lassen sich ableiten aus den §§ 1 und 27 des SGB VIII:

1. Junge Menschen sollen in ihrer individuellen und sozialen Entwicklung gefördert werden,
2. Benachteiligungen sollen vermieden oder abgebaut werden,
3. Eltern und andere Erziehungsberechtigte sollen bei der Erziehung beraten und unterstützt werden,
4. Kinder und Jugendliche sind vor Gefahren für ihr Wohl zu schützen,
5. es sollen positive Lebensbedingungen für junge Menschen und ihre Familien sowie eine kinder- und familienfreundliche Umwelt erhalten oder geschaffen werden.

Auf den unterschiedlichen Steuerungsebenen kommen gegebenenfalls weitere Steuerungsziele hinzu.

2.2 Steuerungsebenen

Man kann sich die Jugendhilfe als ein hierarchisches System vorstellen, in dem es Teilsysteme und Rollendifferenzierung mit unterschiedlichen Aufgaben (auch Steuerungsaufgaben) gibt, die aufeinander aufbauen und zusammenwirken.

1. Die Makroebene:

Die Grundsatzziele der Jugendhilfe hat der Gesetzgeber in Gesetzen und Verordnungen festgelegt. Gleichzeitig hat er den kommunalen Jugendhilfeträgern einen Gestaltungsraum eröffnet und die Landesjugendämter mit Koordinierungsaufgaben betraut. In Berlin gibt es auf der Bezirksebene und auf der Senatsebene jugendhilfopolitische Entscheidungsräume. Die Entscheidungen dienen als Leitlinien und Rahmenbedingungen für die Jugendhilfe. Die Makroebene versorgt die Mesoebene mit strategischen Zielen, strukturellen Rahmungen und mit nötigen Ressourcen.

Die Makroebene braucht Systeminformationen auf einer allgemeinen und aggregierten Ebene. Ein Zuviel an Differenzierung und die Abbildung der gesamten Feldkomplexität führt zu einem „Informations-Overkill“ und führt in eine Komplexitätsfalle (Pruckner, 2005). Zuviel Vereinfachung ist auch schädlich, weil sichergestellt sein muss, dass die Realitäten und Aufgaben der Mesoebene in die Entscheidungsfindung aufgenommen werden. Die Makroebene hat in einem Qualitätsmanagement über die Formulierung von Zielen hinaus auch immer die Aufgabe, die beauftragten Strukturen instandzusetzen (empowerment), ihre Aufgaben erfolgreich umsetzen zu können.

Wirkungsorientierung bedeutet hier, dass sich Leitung nicht nur um die Zielsetzung kümmert, sondern auch überprüft, ob und in welchem Maße Ziele erreicht werden. Wirkungsevaluation bei politischen Programmatiken und Programmen werden zwar immer wieder gefordert, werden aber bisher selten verbindlich durchgeführt. Auf der Makroebene wäre z. B. denkbar, einmal im Jahr einen Wirkungsbericht im Abgeordnetenhaus zu diskutieren, Rahmensexplikationen auf ihre Sinnhaftigkeit zu überprüfen und Entscheidungen nicht nur auf Mengen- und Finanzkennzahlen aufzubauen, sondern Bedarfsentwicklungen und die Leistungsfähigkeit des Jugendhilfesystems in den Blick zu nehmen. Ein anderes Gremium, das Wirkungsanalysen mit bestimmten Schwerpunktsetzungen nutzt, könnte der Rat der Bezirksbürgermeister sein.

Wirkungsberichte auf dieser Ebene entstehen aus der Aggregation der Daten aus den Jugendämtern. Die Unterschiede zwischen den Bezirken können ebenfalls bei den Bewertungen herangezogen werden, wenn dieses in einem lernfreundlichen Klima der Fehlerfreiheit geschieht. Wenn Unterschiede zu einem Ranking zur Ermittlung der besten und schlechtesten Plätze genutzt werden, ist zu befürchten, dass die Wirkungsberichte nicht mehr die Wirklichkeit widerspiegeln, sondern strategisch aufbereitet werden.

2. Die Mesoebene:

Diese umfasst die Organisationen, die politische Ziele umsetzen. Dieses im Fall der Jugendhilfe die Jugendämter und Landesjugendämter. Zu Umsetzung von Zielen richten die Ämter ihr Verhaltungshandeln an diesen aus und koordinieren die Schnittstellen zu den Leistungserbringern, Wohlfahrtsverbänden und Fachverbänden.

Die Vielfalt der Jugendhilfeorganisationen in Berlin ist sehr groß. Wenn man hier nicht nur Jugendhilfe-Institutionen, sondern auch Netzwerke mit Schnittstellen zu anderen Bereichen (Schulen, Tagesstätten, Gesundheitsbereich, Polizei und Justiz usw.) in die Betrachtung einbezieht, kommt man sehr schnell zu dem Ergebnis, dass dieses System nicht mehr top-down steuerbar ist, sondern dass sich die Jugendhilfelandchaft nur evolutionär aus sich heraus gestalten kann.

Wie alle Institutionen auf der Mesoebene funktioniert auch das Jugendhilfesystem wie ein soziales lebendiges System, das in jeweils eigenen Umwelten versucht zu überleben, sich zu entwickeln und einen Beitrag zu einem als wertvoll erachteten Ganzen zu liefern.

Um das Jugendhilfesystem zu einem abgestimmten Miteinander zu bringen, gibt es Gremien, Verträge und Vereinbarungen (z. B. die Leistungsverträge), institutionalisierte Dialoge, gemeinsame Fortbildungen und Fachtag, Projekte (wie z. B. das hier beschriebene) und viele informelle Netzwerke. All diese geschieht zur

Qualitätsentwicklung der Jugendhilfe. SGB VIII § 79a weist dem Jugendamt die Verantwortung dafür zu. Die Art und Weise unterliegt hier wieder der freien Gestaltung.

Zwei Erfordernisse sind gesetzlich verankert: Die Notwendigkeit, dass ein Konsens bezüglich der Qualitätskriterien (Maßstäbe, an denen Qualität zu messen ist) herbeizuführen ist, und zweitens, dass Verfahren zur Sicherung dieser Qualitätserfordernisse umgesetzt werden müssen. Insofern hilft die Implementierung eines einheitlichen Evaluationsstandards in verbindliche Planungs-, Steuerungs- und Prüfprozesse (Fachcontrolling) dem Inhalt und dem Sinn des § 79a SGB VIII in Berlin gerecht zu werden.

Steuerungsaktivitäten auf der Mesoebene beziehen sich im Wesentlichen auf die Prozesse in den Organisationen, auf die Organisation des Zusammenspiels mehrerer Organisationsteile und mehrerer Partner und auf das Empowerment der MitarbeiterInnen, damit das Geplante auf der Mikroebene umgesetzt werden kann.

3. Die Mikroebene:

Hilfen zur Erziehung realisieren sich in Tausenden von Fallverläufen. Hier erweist sich letztendlich, ob die Bemühungen auf der Makro- und der Mesoebene die gewünschten Ergebnisse erbringen. Im Qualitätsmanagement spricht man von den Primärprozessen, also den eigentlichen, wertschöpfenden Handlungen des Jugendhilfesystems. Steuerung auf dieser Ebene bedeutet, Interaktionen so zu gestalten, dass Erwartungen der Klienten und der Gesellschaft (z. B. beim Kinderschutz oder bei Fragen der öffentlichen Sicherheit) befriedigt werden. Steuerung bedeutet hier im Wesentlichen professionelle Selbststeuerung, Steuerung durch Teamreflexion und durch das Fachcontrolling der Führungskräfte.

Die wirkungsorientierte Hilfeplanung umfasst die professionellen Handlungen, mit denen notwendige und geeignete Interventionen geplant, gesteuert und überprüft werden. Der Schwerpunkt liegt auf dem Wort „geeignet“. Der Erfolg einer Hilfe zeigt sich nicht in regelkonformer Bearbeitung der Ablaufschritte und dem korrekten Ausfüllen von Formularen. Auch ist nicht die Auswahl und Genehmigung einer Sozialleistung das Ziel. Die Ergebnisqualität zeigt erst dann, wenn die Hilfe geeignet war, ein Problem zu lösen und ein gesetztes Ziel zu erreichen.

Wirkungsevaluation bedeutet auf der Ebene der Fallbearbeitung, Wirkungen oder ausbleibende Wirkungen festzustellen und zu dokumentieren. Wirksamkeit der Jugendhilfe besteht aus den Atomen der Einzelwirkungen. Da diese im Einzelfall vorhanden sind oder nicht, fällt es Praktikern schwer, in Wahrscheinlichkeiten zu denken. Dennoch müssen sie zu Beginn einer Hilfe eine Entscheidung auf Prognosen aufbauen, die in Form von Wahrscheinlichkeiten (Chancen des Gelingens, Risiken des Misslingens) aufzubauen. Erfahrungswissen, Intuition oder möglicherweise evidenzbasierte Kennzahlen müssen abgewogen, reflektiert, überprüft werden und in den Dialog mit Klienten und in Entscheidungen einmünden. Wirkungsorientierte Steuerung im Case-Management bedeutet, in Hinblick auf die Ziele die wahrscheinlich erfolgreichsten Wege einzuschlagen.

Die Dokumentation der Informationen, die für eine Wirkungsevaluation benötigt werden, kann gleichzeitig dazu dienen, das professionelle Bewusstsein wachzuhalten, dass es nicht nur um Zielvereinbarungen geht, sondern darum, die Wahrscheinlichkeit einer Zielerreichung zu erhöhen und auf dem Weg vom Problem zur Lösung den roten Faden in der Hand zu halten.

4. Die Verschränkung der Steuerungsebenen

Im ungünstigen Fall werden zwar die Aufgaben innerhalb jeder Ebene ordentlich erfüllt (die Makroebene plant, die Mesoebene organisiert, die Mikroebene bewältigt den Alltag), aber jede Ebene bleibt isoliert mit sich selbst beschäftigt. In diesem Fall kommen Pläne nicht in der alltäglichen Wirklichkeit an, die Planenden haben keine Ahnung von der Welt da draußen und der Organisationsapparat perfektioniert sich zu einem bürokratischen Monster, das Politik zu bändigen versucht und das die Mitarbeiter möglichst umgehen.

Systeme, die Übergänge schaffen können, insbesondere entlang der hierarchischen Achse, sind eindeutig erfolgreicher. Dazu bedarf es eines Informationsflusses von unten nach oben und umgekehrt. Allgemeine Richtlinien von oben werden in Verfahren und Organisationsabläufe übersetzt, Ressourcen werden an den richtigen Ort zum richtigen Zeitpunkt gebracht, Mitarbeiter werden auf die Arbeit vorbereitet, Organisationsstrukturen auf die Aufgaben angepasst, passende Partnerschaften eingeleitet und gepflegt. Im beruflichen Alltag werden

die Ziele noch einmal konkretisiert und in methodisches Handeln umgesetzt. „Catch the ball“ nennt man dieses Organisationsprinzip. In der umgekehrten Richtung wird der Ball nach oben gereicht: Informationen, Zielerreichungen, Mängelhinweise werden nach oben gereicht, so dass das Management und die oberste Leitung immer ein Bild von den tatsächlichen Vorgängen in dem System haben.

Nachfolgend wird dieser Zusammenhang an dem Wirkungsevaluationsprojekt erläutert.

Von oben nach unten:

1. Politik möchte wissen, welche erzieherischen und anderen sozialen Bedarfe es in der Bevölkerung gibt, wie sich gesellschaftliche und ökonomische Entwicklungen darauf auswirken, und wie das bestehende System mit den gegenwärtigen Ressourcen Wirkungen erzeugt im Sinne der gesetzten politischen Ziele.
2. Die beauftragten Organisationen erarbeiten Kennzahlensysteme und stellen organisatorisch sicher, dass Informationen von der Basis eintreffen. Zugleich erklären sie den Mitarbeitenden die Sinnhaftigkeit und die Chancen, die in einer wirkungsorientierten Vorgehensweise und Dokumentation liegen, und sie versehen die Mitarbeiter mit Methoden und Hilfsmitteln, Zeitressourcen und Aufmerksamkeit.

Von unten nach oben:

3. Die MitarbeiterInnen und deren Kooperationspartner strukturieren einen Teil ihrer Arbeit neu, nutzen selber direkt einen Teil der neuen Informationen und reichen Daten nach oben weiter.
4. Die Organisation bereitet die Daten auf, aggregiert sie und stellt sie in Zusammenhänge. Einen Teil der Informationen nutzt sie selber, um ihre Abläufe, Regelwerke und Organisationsprinzipien zu verbessern. Einen Teil, der strategisch interessant und verwertbar ist, reicht sie auf die politische Ebene hinauf, um damit die Fragen zu beantworten, die von dort gestellt wurden.
5. Die politische Ebene nutzt die Kennzahlen, um Zielerreichung zu überprüfen, Ressourcen und Rahmenbedingungen nachzubessern, neue Ziele auszumachen, neue Fragen zu generieren, um so den Kreislauf neu zu starten.

2.3 Kennzahlen als Instrument für Steuerung

Wer ein System steuern will, muss klare Vorstellungen von den gewünschten Zuständen und den tatsächlichen Ist-Zuständen haben. Diese Information muss nicht notwendigerweise über Messzahlen erfolgen, allerdings haben quantifizierende Messinstrumente einige Vorteile gegenüber qualitativen, narrativen und in der Regel auch sehr zufällig ausgewählten und subjektiv gefärbten Informationen. Kennzahlen, die auf Messoperationen beruhen, zwingen zu genauen Definitionen und zu einer genauen Verständigung über Maßstäbe. Messen ist letztlich die Abbildung von Merkmalsausprägungen eines Objektes oder eines Vorganges in den Zahlenraum nach definierten Regeln.

Obwohl viele Skeptiker behaupten, man könne in der sozialen Arbeit eigentlich nichts messen, lassen sich leicht sowohl theoretische als auch praktische Gegenbeispiele finden. Die empirische Sozialwissenschaft misst bereits seit weit über 100 Jahren weiche Daten. Es kommt hier im Wesentlichen darauf an, sich auf Definitionen und Maßstäbe zu einigen. Das Problem der Jugendhilfe als professionelles Feld ist es, dass man sich nicht auf Maßstäbe hat einigen können. Im Gegenteil, die Jugendhilfe ist eher stolz darauf, dass es eine große Vielfalt und ein weites Feld lokaler Variationen gibt. Dieser Umstand macht es allerdings schwer, generelle Aussagen über die Jugendhilfe zu machen oder eine Vergleichbarkeit zwischen unterschiedlichen Systemen herzustellen.

In der Untersuchung wurden definierte Skalen eingesetzt, mit denen Bedarfe, Problembelastungen und Zufriedenheit mit Ergebnissen hinlänglich genau gemessen werden können. Dabei streben wir mindestens eine Rangskala an, so dass Aussagen darüber gemacht werden können, bei welchen Fällen wir eine mehr oder weniger hohe Problembelastung vorfinden. Mit etwas Geschick kann man sogar Intervallskalen definieren, wenn zum Beispiel die einzelnen Skalenstufen so definiert sind, dass man in etwa gleichweite Abstände zwischen den unterschiedlichen Skalenpunkten hat. Dieses lässt sich später statistisch überprüfen und hat bisher durchaus zu funktionierenden Kennzahlen geführt.

Kennzahlen sind nicht mit den Messzahlen (Skalierungen) gleichzusetzen, sondern erfordern eine statistische Bearbeitung der Rohdaten. Möglicherweise braucht es mehrere Einzelwerte, um die Zuverlässigkeit (Reliabilität) von Messwerten zu erhöhen. Kennzahlen können berechnet werden mittels Aggregierung mehrerer Einzelbeobachtungen oder Einzelurteile. Manchmal werden auch verschiedene Messwerte in Beziehung gesetzt.

3 Messung von Wirkung und Wirksamkeit - am Beispiel der WIMES-Methode

3.1 Unterschiedliche Evaluationsverfahren

Wirkung und Wirksamkeit können auf unterschiedliche Weise definiert, gemessen und statistisch aufbereitet werden. Bei der Auswahl eines Verfahrens muss man entscheiden, wer die Messung vornimmt, wann Daten erhoben werden, wie viele Informationen man braucht und wie man zu einem Beurteilungsmaßstab kommt.

Beurteiler:

Einig ist man sich darin, dass die Wirkung in irgendeiner Weise die Zielerreichung der Hilfe abbilden muss. Es gibt Wissenschaftler, die einen radikal kundenorientierten Standpunkt einnehmen und von jeglicher fachlicher Einschätzung absehen wollen, sondern nur die Klienten danach fragen wollen, ob sie einen Nutzwert durch die Hilfe erfahren haben (Oelerich, 2005). Davon unterscheidet sich die Wirkungseinschätzung auf fachlicher Sicht. Diese wiederum kann von Akteuren im zu evaluierenden Jugendhilfesystem selber vorgenommen werden (Selbstevaluation) oder von externen Experten (externe Evaluation)

Wir sind zwar durchaus der Meinung, dass die Klienten bei der Bewertung der Ergebnisse eine maßgebliche Stimme haben sollen, glauben aber, dass die Messgenauigkeit bei einer ausschließlichen Nutzerbefragung gering sein wird. Erstens gibt es immer das Problem des mangelnden Rücklaufes, insbesondere von den Klienten, bei denen Dinge nicht so erfolgreich gelaufen sind. Zweitens kann es durchaus sein, dass Klienten in einem erfolgreichen Verlauf überhaupt erst erkannt haben, wie desolat ihre Situation ist, so dass der Erfolg der Hilfe sich in einer deutlich schlechteren Einschätzung der Lebenssituation und der Entwicklung der Kinder widerspiegelt, als dieses am Anfang der Fall war. Eine andere Fehlerquelle ist die, dass im Verlauf einer Veränderung sehr schnell Gewöhnungseffekte eintreten. Dieses Ergebnis gibt es in der Glücksforschung, die bei einem glücklichen Lebensereignis immer nur kurzfristig einen Zuwachs an Lebensqualität und Zufriedenheit feststellt. Danach pendelt sich das System wieder auf den alten Mittelwert ein. Wir befürchten also, dass eine ausschließliche Nutzerbefragung zu sehr schlecht interpretierbaren Kennzahlen führen. Als Ergänzung zu einer fachlichen Einschätzung möchten wir diese dennoch empfehlen.

Für Fremdevaluationen spricht einiges. Wissenschaftliche externe Bewertungen verfolgen in der Regel, was die Ergebnisse angeht, kein Eigeninteresse. Insofern ist mit einer hohen Objektivität zu rechnen, wenngleich auch bei externen Beurteilern immer subjektive Verzerrungen auftreten können. Mittels gut ausgearbeiteter Instrumente und einer guten Schulung von Beobachtern und Interviewern können diese minimiert werden. Externe Evaluationen haben aber auch Nachteile. Der Aufwand ist in der Regel sehr hoch und mit hohen Kosten verbunden. Das Hauptproblem liegt aber in der Akzeptanz der Ergebnisse seitens der Akteure des Jugendhilfesystems und die Übertragbarkeit in die Praxis. Ob aus interner Sicht die Dinge richtig erfasst und beurteilt worden sind, wird oft bezweifelt. Letztlich führt das untersuchte System immer eine interne Plausibilitätskontrolle anhand einer informellen Selbstevaluation durch. Eine weitere Schwierigkeit liegt darin, dass eine externe Evaluation einen direkten Zugang zu den Klienten braucht, der nicht in jedem Fall herzustellen ist und sich eventuell auch als störend und die Hilfe beeinflussend auswirken kann. Explorationen, Interviews, Akteneinsicht, teilnehmende Beobachtungen setzen das Vertrauen der Klienten voraus und einen einfühlsamen Umgang mit ihnen. Das ist alles zu lösen, bedeutet aber einen hohen Aufwand und ist nicht frei von Fehlerquellen.

Auch Selbstevaluationen unterliegen Beurteilungsfehlern. Subjektive Verzerrungen und „blinde Flecken“ beeinflussen die Datenerhebung. Selbstevaluation hat allerdings den Vorteil, dass die Akteure den Daten vertrauen und sie unmittelbar verstehen, was sie bedeuten. Dieses erleichtert den Transfer von Erkenntnissen in die Praxis. Fehlermindernd ist der Einsatz eines standardisierten praxisnahen Messinstrumentes. Unterschiedliche Maßstäbe oder Schwerpunktsetzungen werden homogenisiert, was die Verlässlichkeit und statistische Auswertbarkeit erhöht. Es darf nicht unterschätzt werden, dass auch Selbstevaluation Zeit und Aufwand mit sich bringt. Da die Fachleute aber ohnehin Informationen von Klienten erheben, bedarf es keiner zusätzlichen Zeit für die Informationsgewinnung, sondern lediglich eines gewissen Aufwandes für die Aufbereitung der Informationen und deren Dokumentation. Selbstevaluationen können gegen allzu große Fehlereinflüsse geschützt werden durch eine externe wissenschaftliche Begleitung. Diese kann geeignete erprobte Messverfahren entwickeln oder zur Verfügung stellen, die Qualität der Daten überwachen und statistische Berechnungen durchführen.

Die Erhebungszeitpunkte

Bei der retrospektiven Evaluation werden die Werte am Schluss der Hilfe rückblickend erhoben. Der Beurteiler muss sich daran erinnern, was das Problem und das vereinbarte Ziel war. Außerdem muss er sich an die Maßstäbe der Problembewertung erinnern, die er am Anfang angewendet hat. Retrospektive Einschätzungen sind äußerst fehleranfällig, weil Gewöhnungseffekte und nachträgliche Verklärungen oder Dramatisierungen die Urteile stark beeinflussen. Nach unseren Erfahrungen wirken sich bei der Urteilsbildung oft die letzten Ereignisse auf das Letzturteil aus. Wenn Familienhilfe beispielsweise in einer Krise endet oder in den letzten Wochen eine Verschlechterung eingetreten ist, gerät häufig der anfängliche Erfolg völlig aus dem Blick. Zum anderen ist nicht immer sichergestellt, dass bei einer rückblickenden Bewertung die anfänglichen Ziele noch genau erinnert werden. Es würde natürlich helfen, wenn die Ziele am Anfang genau operationalisiert wurden und am Ende retrospektiv mit dem Wissen dieser anfänglichen Beauftragung eine Bewertung vorgenommen wird. In der Praxis können wir aber nicht davon ausgehen, dass am Schluss die anfänglichen Ziele immer noch aktuell sind, und wir können auch nicht davon ausgehen, dass die Zielvereinbarungen am Beginn so genau getroffen worden sind, dass sie am Ende einer Hilfe noch Bestand haben.

Prospektive Evaluationen vermeiden die meisten Fehler der retrospektiven Wirkungseinschätzung. Die Daten werden zu Beginn, im Verlauf und am Ende der zu beurteilenden Intervention erhoben. Durch die zeitnahe Dokumentation vermeidet man Erinnerungsverzerrungen. Die Erhebung im Verlauf der Hilfen setzt aber voraus, dass die Maßstäbe über die Zeit stabil sind. Sollten diese sich mit der Zeit verschieben, weiß man am Schluss nicht, ob Veränderungen tatsächlich die Wirkungen von Interventionen widerspiegeln oder auf Verschiebungen der Maßstäbe. Prospektive Evaluationen sind wissenschaftlich zu empfehlen, haben aber den Nachteil einer langen Datenerhebungsphase.

Einen guten Kompromiss bieten quasi-experimentelle Querschnittuntersuchungen. Wie bei der prospektiven Evaluation werden mindestens zwei Zeitpunkte dokumentiert, hier der Hilfebeginn und das Hilfeende. Es bedarf auch hier eines einheitlichen Beurteilungsmaßstabes. Es ist aber nicht nötig, jede Hilfe vom Anfang bis zum Ende zu verfolgen, indem man zwei Vergleichsgruppen bildet, in unserem Fall die Stichprobe der begonnenen in einem Zeitraum im Vergleich zu den beendeten Hilfen des Zeitraumes. Dieses Verfahren ist zeitökonomischer und eröffnet bei längerer Dauer die Möglichkeiten einer echten prospektiven Längsschnittuntersuchung. Die Stichprobe der beendeten Hilfen kann außerdem noch retrospektiv evaluiert werden.

Der Detaillierungsgrad:

Die Ziele der Intervention können sehr differenziert erfasst werden oder es wird ein eher generalisierter Blick auf die Ziele und deren Erreichung geworfen. Ein Beispiel für eine stark differenzierte Ziel-Wirkungsevaluation bei Erziehungshilfen ist die Dokumentation von smarten Handlungszielen. Diese können kaum kategorisiert werden, weil es eine nicht begrenzte Anzahl smarter Zielvereinbarungen geben kann. Der Nachteil stark differenzierter Ziele liegt darin, dass der Dokumentationsaufwand sehr hoch ist und die Menge der zusammengetragenen Information schwerlich eine allgemein gültige Aussage erlaubt.

Die allgemeinste Form einer sehr wenig differenzierten Zielerreichung ist eine Ja-Nein Frage: Hat die Intervention gewirkt? Mit einer solchen allgemeinen Einschätzung kann Praxis in der Regel wenig anfangen. Abgesehen davon, dass auch von Interesse ist, worin denn die Wirkung besteht, sollte man berücksichtigen, dass Ziele in abgestuften Graden erreicht werden.

Dazwischen liegt ein ganzes Spektrum von Detaillierungsgraden, auf dem man je nach Fragestellung und möglichem Aufwand die geeignete Form finden muss. Je differenzierter, desto höher der Aufwand und die zu verarbeitende Informationsmenge. Je weniger differenziert, desto geringer der Erkenntnisgewinn.

Das Problem der Vergleichsgruppe.

Die Überprüfung von Wirksamkeit einer Intervention setzt voraus, dass man eine Vergleichsgruppe definiert, bei der die Intervention nicht angewandt wurde (Schmidt, 2010). Dieser sogenannte „Goldstandard“ der Wirkungsevaluation kann allerdings in der Jugendhilfe in der Regel nicht angewandt werden, weil es aus ethischen und praktischen Gründen wahrscheinlich nicht gelingen wird, eine Gruppe zu bilden und im Verlauf der nächsten Jahre nachzuverfolgen, bei der ein entsprechender Bedarf existiert und keine Hilfe gewährt wurde. Falls eine solche Vergleichsstichprobe gefunden werden kann, stellt sich immer noch die Frage, ob es sich

tatsächlich um eine merkmalsgleiche Gruppe handelt, oder nicht doch um eine Gruppe, die zum Beispiel von Jugendhilfe nicht erreichbar ist und nicht motiviert ist, eine Hilfe anzunehmen. Zum anderen dürfte es auf praktische Schwierigkeiten stoßen, diese Gruppe, die sich nicht auf eine Hilfe einlassen will, zu einer Teilnahme an einem Forschungsprojekt zu bewegen. Nach einem randomisierenden Verfahren einem Teil der Klientel aus wissenschaftlichen Gründen eine Hilfe vorzuenthalten, dürfte sich aus rechtlichen und ethischen Gründen verbieten.

Jugendhilfeevaluation wird sich in der Regel mit einem geringeren Anspruch zufrieden geben müssen, was deswegen nicht schlimm ist, weil meisten die zentrale Fragestellung nicht ist, ob Erziehungshilfe wirkt, sondern in welcher Weise welche Formen bei welchen Problemlagen in welchem Ausmaß Wirkungen zeigen.

Als Ersatz für eine Non-Treatment-Gruppe wählen kann man einen Benchmark-Vergleich machen und überprüfen, ob eine vergleichbar bedürftige Klientel in anderen Settings wirksamer mit Hilfen zur Erziehung versorgt wird oder nicht. Solche Benchmark-Vergleiche klären nicht die Frage, ob die Zeit ohne professionelle Hilfe nicht auch Verbesserungen gebracht hätte. Trotzdem kann es hier einen Anhalt geben, ob die ermittelte Wirksamkeit kleiner ist als in anderen Settings.

Einen anderen Vergleichspunkt kann man auch in den Ausgangsbedingungen sehen. Man könnte beispielsweise die Gruppe der Klienten, die am Anfang der Hilfe stehen, als eine Kontrollgruppe ansehen, die eben bisher noch nicht die zu evaluierende Hilfe bekommen hat. Hier haben wir lediglich noch eine kleine methodische Schwierigkeit, die darin liegt, dass diese Klientel jünger ist. Letztlich lässt sich aber ein Wirkungsnachweis mit etwas gutem Willen dadurch erbringen, dass man feststellt, dass Klienten, die vor der Hilfe möglicherweise jahrelang in Schwierigkeiten waren, mit Hilfe einer Intervention aus diesen Schwierigkeiten herausgekommen sind.

3.2 Wirkungsmessung mit der WIMES-Methode

Für die Untersuchung der Wirksamkeit in den 5 Projektbezirken wurde das WIMES-Verfahren² (Tornow, 2012) eingesetzt. Nach den obigen Kriterien, ist WIMES

- a) ein wissenschaftlich kontrolliertes Selbstevaluationsverfahren
- b) das in einem prospektiven Evaluationsverfahren (quasi-experimentelles Design mit der Möglichkeit einer Längsschnittuntersuchung)
- c) mit retrospektiven Ergänzungen
- d) auf einem mittleren Detaillierungsgrad
- e) Wirkungskennzahlen ermittelt,
- f) die zu den Ergebnissen einer Benchmarking-Stichprobe in Beziehung gesetzt werden.

1. **Wirkung** wird definiert als eine erwünschte Veränderung problematischer Zustände oder Entwicklungen, die mit hinreichender Plausibilität auf eine Intervention (hier: HzE) zurückzuführen ist. Bei Hilfen zur Erziehung geht es um erzieherische Bedarfe, Gefährdungen, Kinderrechte und Teilhabechancen. Wirkungen sind der Zugewinn an Kompetenzen, Ressourcen, Chancengerechtigkeit und Teilhabe.

2. **Wirksamkeit** (Effektivität) ist das Ausmaß erzielter Wirkungen, gemessen an definierten Maßstäben (Vereinbarungen, Benchmarks, Zeitreihen usw.). Kennzahlen für die Effektivität sind die durchschnittlichen Mittelwertunterschiede zwischen Anfang und Ende oder die relative Häufigkeit der positiv verlaufenen Fälle oder andere retrospektive Einschätzungen.

² Dieses Verfahren ist seit 2006 in ca. 150 Einrichtungen und einigen Kommunen im Einsatz. Zurzeit liegt eine Datenbank mit ca. 50.000 Datensätzen vor.

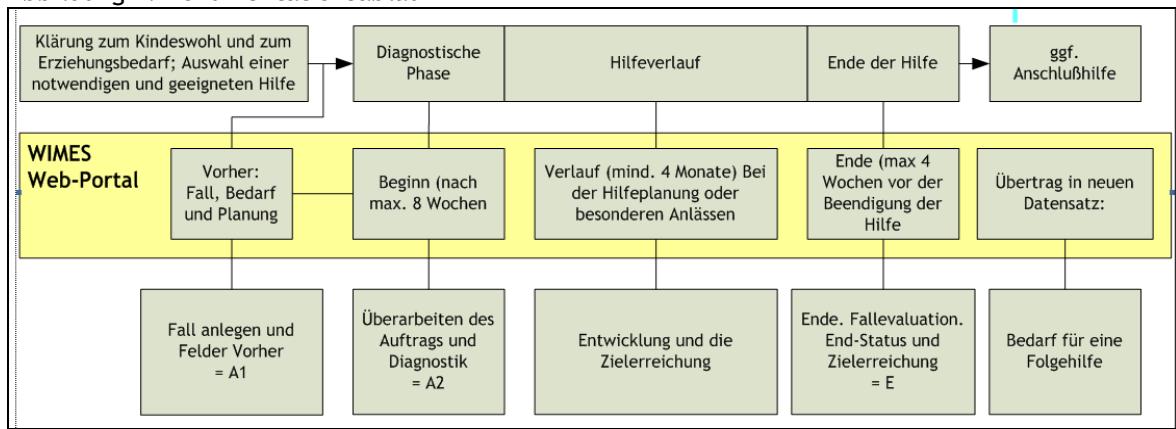
3. Es gibt drei Messzeitpunkte:

A1: Vor Beginn der Hilfe werden der erzieherische Bedarf, die Grundsatzziele und die Richtungsziele dokumentiert.

A2: Nach ungefähr 6 bis 10 Wochen werden die diagnostischen Einschätzungen und die Grundsatz- und Richtungsziele überprüft und dokumentiert.

E: Am Ende der Hilfe werden - wie am Anfang - Einschätzungen der Ressourcen / Kompetenzen versus Problembelastungen und Risiken vorgenommen. Zusätzlich wird die Hilfe retrospektiv bewertet.

Abbildung 1: Dokumentationsablauf

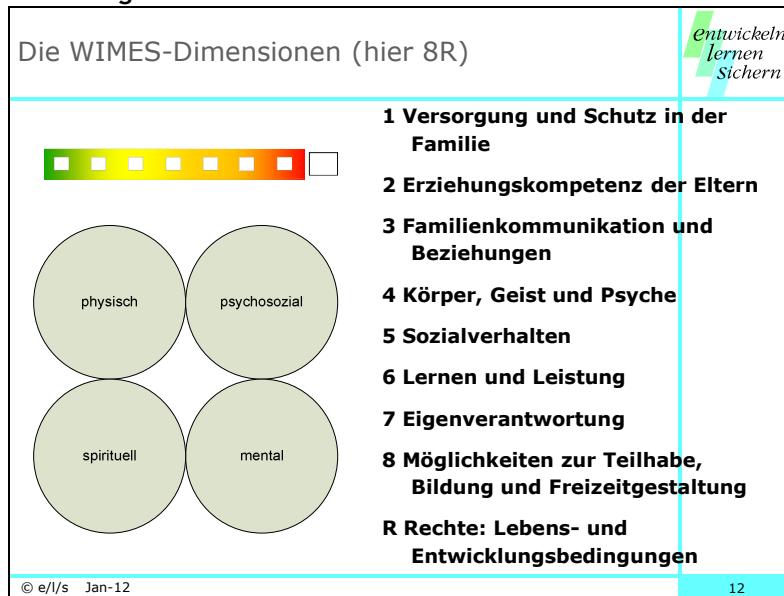


4. Die Einschätzung der Bedarfe und andere Bewertungen erfolgen nach definierten Standards. Dafür wurden den Mitarbeitern und Mitarbeiterinnen in den sozialen Diensten Werkzeuge an die Hand gegeben (siehe Anlage 7 und das Verfahrensmanual). Alle MitarbeiterInnen wurden in dem Verfahren geschult und werden durch eine Hotline des e/l/s-Instituts unterstützt. Für die Koordination vor Ort wurden Personen benannt.

Die Dateneingabe erfolgt elektronisch. In dem Eingabeprogramm wird durch Eingabehilfen, Terminerinnerungen und Plausibilitätsprüfungen die Einhaltung des Messverfahrens unterstützt.

5. Die neun WIMES-Dimensionen beziehen sich auf die Familie, den jungen Menschen und sein Umfeld (Abbildung 2). Die Bedarfssausprägungen werden anhand sogenannter „Ankerbeschreibungen“ skaliert.

Abbildung 2: Bedarfsdimensionen nach WIMES



Die Auswahl der neun³ verwendeten Dimensionen basiert auf den gesetzlichen Vorgaben (erzieherische Bedarfe im Sinne des SGB VIII), auf der Hilfeplanpraxis (die weitaus meisten Hilfeziele sind den Dimensionen zuzuordnen) und auf dem Vergleich zu anderen Katalogen (z.B. Capability-Approach (Otto & Ziegler, o.J.)(Otto, 2010)).

6. Als Untersuchungsdesign wurde eine quasi-experimentelle Querschnittuntersuchung gewählt (Abbildung 3). Diese hat gegenüber einer Längsschnittuntersuchung den Vorteil, dass relativ zeitnah Wirksamkeitsschätzungen vorgenommen werden können. Längsschnittuntersuchungen brauchen mindesten zwei bis drei Jahre, um genug abgeschlossene Fälle in der Stichprobe zu haben. Kürzere Längsschnittuntersuchungen führen zur Negativ-Verfälschung, weil nur Hilfen mit kurzer Dauer erfasst werden, die wegen des höheren Abbruchanteils durchschnittlich weniger erfolgreich sind. Nach ungefähr drei Jahren Dateneingabe sind auch Längsschnittauswertungen möglich.

In dem definierten Zeitraum werden alle Neufälle (für A1 und A2) und alle beendeten Fälle (E) erfasst, egal ob eine Eingangseinschätzung vorliegt oder nicht, wie z. B. bei Altfällen, die vor Projektbeginn begonnen haben. Die Berechnung der Effekte ergibt sich aus dem Vergleich der Bedarfe am Anfang mit den Bedarfen am Ende. Solche Berechnungen unterscheiden sich nach bisherigen Erfahrungen kaum von den Ergebnissen von Längsschnittstudien. Man kann allerdings nur Mittelwertvergleiche anstellen und nicht im Einzelnen erkennen, in wie vielen Fällen es Verschlechterungen oder Verbesserungen gegeben hat. Diese Analysen sind später in Längsschnittauswertungen möglich.

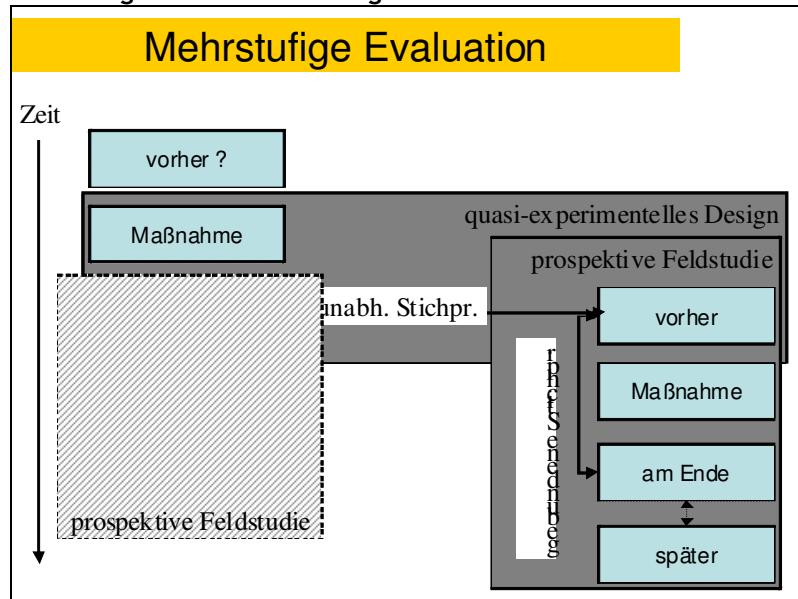
7. Außer deskriptiven Statistiken werden die Effektstärke nach COHEN (Bortz & Döring, 2002) und einige Zusammenhänge berechnet.

Soweit es möglich ist, werden multivariate Verfahren zur Erklärung durch Wirkfaktoren (Wirkursachen) berechnet. Hier dürfen die Erwartungen nicht zu hoch sein, weil es sich bei dem Projekt nicht um „Wirkungsforschung“ handelt. Die Erforschung der Frage, warum eine Intervention wirksam oder unwirksam ist, setzt

³ Aus den Berliner Erfahrungen wurden 2012 einige Veränderungen in dem WIMES-Katalog vorgenommen (Erweiterung durch 3 familienbezogene und sozialräumliche Dimensionen). Diese mit den Bezirken abgestimmte Veränderung ist aber in Berlin nicht mehr zum Tragen gekommen, weil zwischenzeitlich die Datenerhebung beendet wurde.

Hypothesen und Modelle und eine differenzierte Fragestellung sowie eine umfangreichere Datenerhebung voraus. In dem Projekt geht es vorerst nur um die Frage, ob und in welchem Ausmaß Wirkungen erzielt werden und ob sich daraus ein Steuerungsbedarf ergibt. Als Beispiel für eine Wirkungsforschung auf Grundlage der Wirkungsevaluation haben wir die Ursachen von Abbrüchen stationärer Hilfen untersucht.

Abbildung 3: Evaluationsdesign



8. Um die ermittelten Kennzahlen bewerten zu können, bietet WIMES einen Benchmark-Vergleich an. Für den gleich Zeitraum, auf dem die Projekt-Kennzahlen beruhen, wurden nach den gleichen Mess- und Analyseverfahren in Einrichtungen und Jugendämtern der gesamten Bundesrepublik Daten erhoben und Kennzahlen berechnet. Bei der Darstellung der Ergebnisse werden diese verglichen, wenn es sinnvoll erscheint. Bei der Gesamtauswertung der Daten werden Berliner Ergebnisse und Ergebnisse der Benchmark-Stichprobe gegenüber gestellt.

Da die Berliner Stichprobe einen nicht unerheblichen Anteil bei der WIMES-Gesamtdatenbank darstellt, wurden die Berliner Daten aus der Vergleichsstichprobe herausgefiltert. Andernfalls würde Berlin zum Teil mit Berlin verglichen, was keine zusätzliche Information bringt. Es muss allerdings bei den Vergleichen darauf geachtet werden, dass der bundesweite Benchmark auch Daten aus Landkreisen und kleineren Kommunen enthält. Die erhobenen Fälle kommen aus ganz Deutschland, wobei der Süden und Westen Deutschlands überrepräsentiert sind. Rahmenbedingungen, Problemlagen und Angebotsstrukturen sind nur eingeschränkt vergleichbar. Dennoch mag sich aus den Vergleichen eine Möglichkeit der Bewertung der Berliner Ergebnisse eröffnen.

4 Empirische Ergebnisse der Wirkungsevaluation (aggregiert über alle 5 Projektbezirke)

Die Auswertungen der Evaluation können auf unterschiedlichen Aggregationsebenen berechnet werden. Jeder einzelne Datensatz (Fall) ist einem Bezirk und einem Regionalteam zugeordnet. In der Folge werden alle Daten über alle Bezirke hinweg zusammengefasst (aggregiert). Dadurch werden allgemeine Ergebnisse wieder gegeben, die Berlin-weit die Effektivität von Hilfen zur Erziehung beschreiben, jedenfalls soweit es die 5 beteiligten Bezirke betrifft. Unterschiede zwischen den Bezirken werden dadurch nivelliert. Wenn die Unterschiede zwischen den Bezirken hoch sind, wird an den entsprechenden Stellen darauf hingewiesen.

Die differenzierten Berichte über die Bezirksergebnisse werden in gleicher Form aufbereitet und den Bezirken direkt zugestellt. Der Gesamtbericht kann in den Bezirken als interner Benchmark herangezogen werden, um die relative Position der Bezirke innerhalb der fünf Projektbezirke zu bestimmen. Beachtet werden muss jeweils, ob Unterschiede statistisch überhaupt signifikant und inhaltlich relevant sind.

Die Senatsverwaltung und die Bezirken bekommen differenzierte Auswertungen auf einer CD-ROM ausgeliefert (Anhang 9). Die CD enthält auch die gesamten Rohdaten in Tabellenform. Auf diese umfangreiche Sammlung von statistischen Auswertungen muss verwiesen werden, wenn Kennzahlen im Detail interessieren. Diese werden aus nachvollziehbaren Gründen nicht veröffentlicht. In der Folge werden nur die wichtigsten Ergebnisse, die von allgemeinem Interesse sind, im Überblick dargestellt. Sollte es Fragestellungen geben, die noch nicht ausgewertet worden sind und die mit dem vorliegenden Datenmaterial sinnvoll zu beantworten sind, kann das e/l/s-Institut entsprechende Auswertungen auf Anfrage von Senatsverwaltung und Bezirken nachliefern.

4.1 Datenbasis

Die Auswertungen basieren auf einer hinreichend repräsentativen Stichprobe aus allen Hilfen zur Erziehung in den fünf beteiligten Bezirken. Repräsentativ ist eine Teilstichprobe dann, wenn sie eine Zufallsauswahl aus der Gesamtheit aller Fälle darstellt. Aus verschiedenen Gründen, die an späterer Stelle dargelegt werden, ist eine Gesamterfassung nicht gelungen. In diesem Fall würde sich die Frage nach der Repräsentativität nicht stellen. Alle Koordinatoren des Projektes haben versichert, dass es keine gezielte Auswahl gegeben habe, nach der entschieden wurde, ob ein Fall in die Evaluationsstichprobe kommt oder nicht. Wir gehen also davon aus, dass eine Aussage über die Stichprobe der beste Schätzwert für die Kennzahlen der Gesamtheit aller Fälle ist.

Gegen die reine Zufallsauswahl spricht allerdings die Tatsache, dass es einen Zusammenhang zwischen der Wahrscheinlichkeit, dass ein Fall in die Evaluationsstichprobe aufgenommen wurde, und der Hilfeart gab. Während 60% aller ambulanten und teilstationären Hilfen erfasst wurden, waren es nur 34% der stationären Fälle. Dieser Unterschied ist statistisch hoch signifikant (Chi-Quadrat: 1,70499E-67; p < 0,001). Diese Abweichungen gleichen wir durch Gewichtungen der einzelnen Leistungstypen in den Bezirken aus. Insofern ist die Unterrepräsentanz der stationären Fälle nicht gravierend. Schädlicher für die Validität und nicht zu korrigieren wären Selektionsverzerrungen, die etwas mit der Effektivität der Hilfen zu tun haben, indem beispielsweise nicht gut verlaufene Hilfen eher weggelassen werden. Hierauf deutet aber nichts hin.

Da zu den Projekt-Bezirken solche mit dem höchsten Fallaufkommen in Berlin gehören, beziehen sich die folgenden Kennzahlen ungefähr auf die Hälfte aller Erziehungshilfen in Berlin in den Jahren 2011 und 2012.

Für Globalaussagen reichen die fast 12.000 Fälle allemal aus. Für Längsschnittanalysen und für Detailfragen (z. B. einzelne Hilfeformen in bestimmten Regionen für eine bestimmte Klientel) stößt die Stichprobe aber an Grenzen. Dennoch handelt es sich um die größte prospektive Wirkungsevaluation in Deutschland.

Um den dokumentierten Datenbestand in der Evaluationsdatenbank (WIMES-Ist) hinsichtlich seiner Vollständigkeit und Repräsentativität bewerten zu können, wird er in Beziehung zu den Fallstatistiken aus Pro-Jug gesetzt. Eine genauere Aufstellung findet sich in der Anlage 8.

Von den begonnenen Hilfen wurden 60 % erfasst. Bei den Beendigungen waren es 53 %. Dieser Unterschied ist zwar statistisch signifikant, inhaltlich aber zu vernachlässigen. Die Erfassung auch von Beendigungen ist von 2011 zu 2012 deutlich gestiegen. Die absolute Differenz zwischen den Beendigungen zu den begonnenen Hil-

fen in Höhe von 6791 - 5081 = 1710 Fällen geht zum größeren Teil auf den Fallanstieg in Berlin zurück. Insgesamt gibt es einen Netto-Anstieg von 2011 auf 2012 in einer Höhe von 18 %. Das heißt, es kommen deutlich mehr Klienten in das System als dass Klienten das System verlassen. In der WIMES-Stichprobe haben wir 33 % mehr Aufnahmen als Entlassungen (15 % sind also dem Weglassen von Fällen bei der Schlussdokumentation geschuldet).

4.2 Statistische Auswertungen (Lesehilfe)

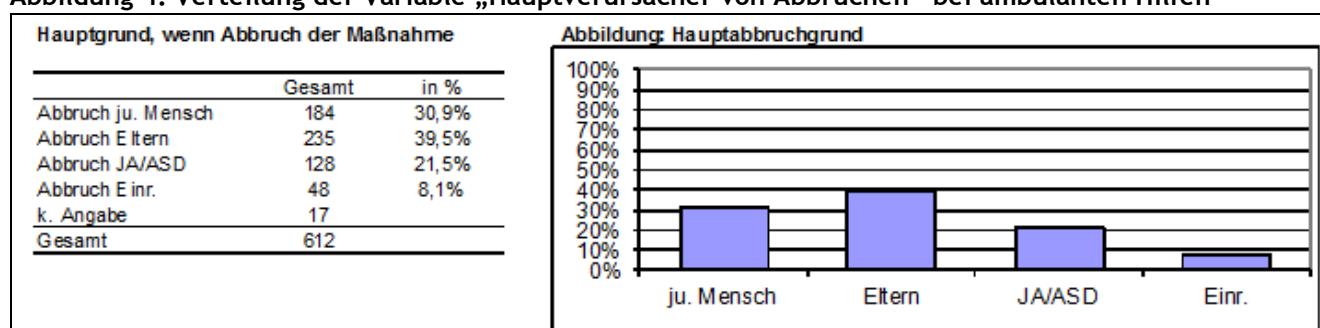
Um die Auswertungen richtig zu verstehen, sollte man einige statistische Grundbegriffe kennen. Eine Übersicht in Form eines Glossars finden Sie in der Tabelle 2 Seite 25.

Zunächst einmal ist zwischen deskriptiven Verteilungskennzahlen und explorativen Auswertungen zu unterscheiden. Bei den deskriptiven Kennzahlen geht es darum, sich einen Überblick über eine größere Menge von Daten zu verschaffen. Explorative statistische Auswertungen haben das Ziel, Zusammenhänge oder Unterschiede zwischen Variablen darzustellen und zu untersuchen, mit welcher Wahrscheinlichkeit diese nicht auf die zufällige Auswahl einer Stichprobe zurückzuführen sind. Letztere Berechnungen sind wichtig, um seine Erklärungen und Prognosen nicht auf Zufällen und falschen Annahmen aufzubauen.

4.2.1 Deskriptive Statistik

Variable sind Eigenschaften, die unterschiedliche Merkmalsausprägungen annehmen können. Man könnte eine Variable als eine Frage definieren, für die es mehrere Antworten gibt, entweder in Form von vorgegebenen Kategorien oder in Form von Messwerten, also Zahlen. Die einfachste Darstellungsweise einer Variablen ist ihre Häufigkeitsverteilung. Dafür wird jeweils ausgezählt, wie oft ein bestimmtes Merkmal auftritt. Die Darstellung kann in Form von Häufigkeitstabellen oder in Form von Grafiken (Histogramme) erfolgen. Zwei Beispiele aus den Wirkungsberichten mögen das verdeutlichen. In Abbildung 4 werden die Häufigkeiten der Hauptgründe für die erfolgten Abbrüche ausgezählt. Um solche Verteilungen für unterschiedliche Stichproben vergleichbar zu machen, sind %-Angaben hilfreich. Dabei muss man darauf achten, dass fehlende Angaben unberücksichtigt bleiben. In unserem Beispiel ist in 17 Fällen die gestellte Frage nicht beantwortet worden. Das ist angesichts der Gesamtzahl von 612 Fällen zu vernachlässigen. Bei %-Angaben sollte man sich auch immer die Absolutzahlen ansehen, um zu überprüfen, ob die Stichprobe groß genug ist.

Abbildung 4: Verteilung der Variable „Hauptverursacher von Abbrüchen“ bei ambulanten Hilfen



Eine andere Verteilung ist in Abbildung 5 dargestellt. Im Gegensatz zu dem obigen Beispiel, in dem Antwortkategorien vorgegeben wurden, sind die einzelnen Jahre in Zahlen ausgedrückt. Bei dieser Verhältnisskala kann die Form der Verteilung (jede Säule steht für ein Quartal) interpretiert werden und es können Verteilungsparameter (siehe unten) berechnet werden. Die Verteilung fällt nach rechts ab, das heißt, die meisten Hilfen enden in den ersten zwei Jahren; lange Hilfzeiträume kommen vor, finden sich aber eher selten.

Wir verwenden noch eine andere Darstellungsart für Häufigkeitsverteilungen: sogenannte kumulierende Balken. Hierfür werden die einzelnen Häufigkeiten quasi aneinander gelegt. Das ist besonders anschaulich bei Ordinalskalen, also Merkmalsausprägungen, die nach der Größe sortiert werden können (Abbildung 6). Je größer der Grünanteil, desto positiver die Entwicklung. Die absoluten und die relativen Häufigkeiten (%-Achse unten) sind in einer einzigen Grafik dargestellt.

Abbildung 5: Verteilung der Variable „Dauer ambulanter Hilfen“

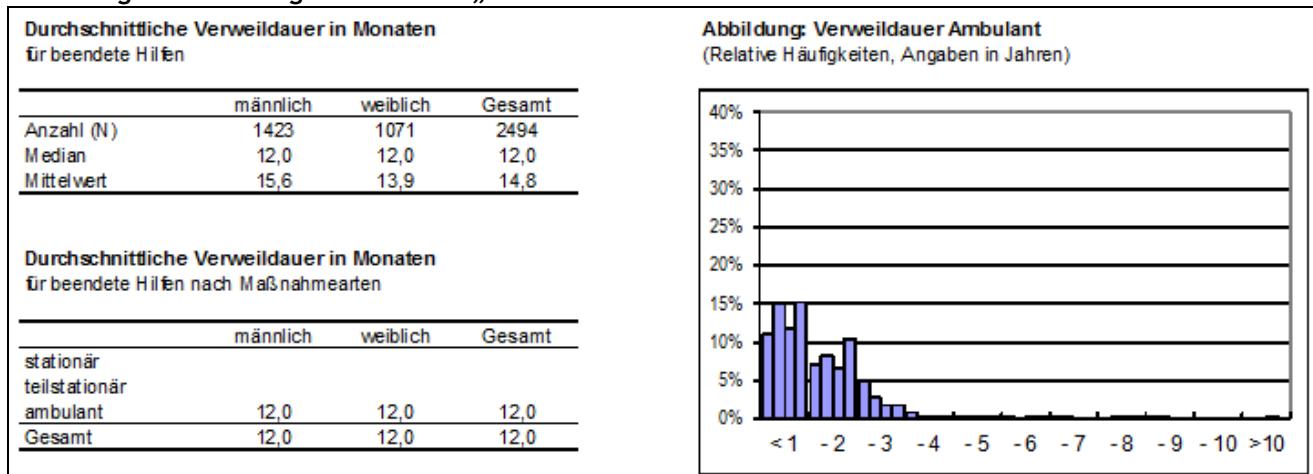
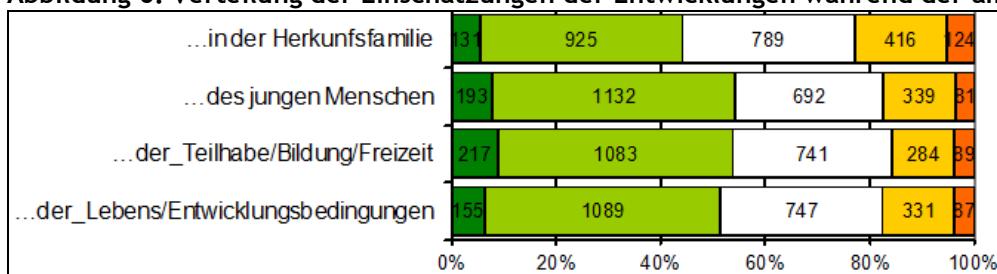


Abbildung 6: Verteilung der Einschätzungen der Entwicklungen während der ambulanten Hilfen



Die Verteilungsparameter erlauben es, die wesentlichen Informationen einer Häufigkeitsverteilung in einer einzigen Kennzahl auszudrücken. Am gebräuchlichsten, weil am informativsten und anschaulichsten ist der arithmetische Mittelwert. Alle vorkommenden Werte werden addiert und durch die Anzahl der Fälle dividiert. Bei der Berechnung eines Mittelwerts muss man allerdings darauf achten, dass die Verteilung ungefähr symmetrisch ist, dass also die Verteilung der abweichenden Werte nach oben und unten ungefähr gleich aussieht. Das ist zum Beispiel in der Abbildung 5 nicht der Fall. Wenige Extremwerte nach oben (hier wenige Hilfen, die sehr lange dauern) ziehen den Mittelwert nach oben. Bei schießen Verteilungen ist es deswegen besser, zur Kennzeichnung der Verteilung den Median anzugeben. Dieser bezeichnet die Merkmalsausprägung, die die Verteilung in zwei gleich starke Hälften verteilt. In unserem Beispieldfall besagt der Median von 12 Monaten, dass 50% der ambulanten Hilfen innerhalb eines Jahres beendet werden. Die andere Hälfte dauert länger.

Ein zweiter gebräuchlicher Parameter ist die Standardabweichung der Verteilung, also die Breite, mit der die Werte um den Mittelwert streuen. Die Standardabweichung brauchen wir hauptsächlich, weil wir damit die Effektstärke berechnen.

Weitere Verteilungsmaße zur Schiefe, zur Verteilungsform, zur Breite der Verteilung usw. werden für das Verständnis der Auswertungen nicht benötigt, weil wir so oft wie möglich die Verteilung selber darstellen.

4.2.2 Explorative Statistik

Mittels statistischer Verfahren werden Unterschiede zwischen Merkmalen, zwischen Fallgruppen und zwischen unterschiedlichen Zeitpunkten untersucht. Die zweite Frage, die geklärt wird, ist die nach Zusammenhängen zwischen Variablen.

Die Darstellung von Mittelwertunterschieden ist relativ einfach: man bildet die numerische Differenz zwischen den Mittelwerten zweier oder mehrere Unterstichproben und prüft, ob diese Differenz wahrscheinlich nicht durch einen Zufall entstanden ist (Signifikanzprüfung).

Da bei der explorativen Statistik die Signifikanz immer wieder eine Rolle spielt, hier dazu eine kurz Erklärung: Bei der Evaluation, wie bei jeder anderen Forschung, interessieren Erkenntnisse, die eine gute Prognose für die Zukunft liefern. Dazu müssen die Ergebnisse wirkliche Gegebenheiten, Strukturen und Muster wiederspiegeln und dürfen nicht auf einmalige, zufällige Ereignisse zurückzuführen sein. Die Wahrscheinlichkeit, dass ein Befund auf einer zufälligen Mischung der Fälle, auf Messfehlern oder auf spezifischen, einmaligen Gegebenheiten beruht, lässt sich berechnen. Es lässt sich niemals vollkommen ausschließen, dass ein Befund auf zufälligen Ereignissen beruht. Wir geben uns damit zufrieden, wenn die Wahrscheinlichkeit kleiner als 5% ist. Das stellt man in dieser Schreibweise dar: $p < 0,05$, die Wahrscheinlichkeit p , dass man einen Unterschied annimmt, wo in Wirklichkeit gar keiner ist, ist kleiner als 5% (Alpha-Fehler). Wahrscheinlichkeiten werden als %-Zahlen oder als Dezimalzahlen dargestellt. In jedem Wirkungsbericht sind also 5% der Ergebnisse Täuschungen. Leider wissen wir nicht, welcher Teilbefund nicht vertrauenswürdig ist.

Die Verlässlichkeit eines statistischen Befundes steigt mit der Zahl der Fälle N und sinkt mit der Streuung der Verteilung. Bei der Auswertung des gesamten Datensatzes aus fünf Bezirken ist die Fallzahl in der Regel so groß, dass der Alpha-Fehler unter 0,1% sinkt. Also sind die Ergebnisse des vorliegenden Gesamtberichts wahrscheinlich alle zuverlässig. Als Faustregel kann man sich merken, dass bei Stichproben unter $N=20$ der Alpha-Fehler stark ansteigt. Man muss also vorsichtig sein mit Vergleichen zwischen einzelnen Regionalteams oder zwischen einzelnen Anbietern, weil dort in der Regel die Fallzahl klein ist. Leider wird hierbei oft nicht die nötige Vorsicht an den Tag gelegt. Erst größere Stichproben oder Evaluationen über zwei bis drei Jahre gestatten verlässlichere Aussagen.

Zurück zu den explorativen statistischen Analysen. Mittelwertvergleiche können nur angestellt werden, wenn ein Merkmal mindestens auf einer Rangskala gemessen werden kann. Wenn eine Variable durch Zuordnung zu Kategorien gemessen wird (Nominalskala), können nur die beiden Verteilungen deskriptiv gegenübergestellt werden. Inwieweit diese sich tatsächlich überzufällig unterscheiden, wird mittels einer Prüfstatistik (Chi-Quadrat) getestet, die man nicht im Detail verstehen muss. Wichtig ist auch hier, dass man darauf achtet, dass der Alpha-Fehler p kleiner als 5% sein sollte.

Wenn man die Stichprobe nach einer bestimmten Variable in Untergruppen teilt (unabhängige Variable) und dann einen Unterschied zwischen den Mittelwerten oder den Häufigkeitsverteilungen (der abhängigen Variable) feststellt, kann das auch so interpretiert werden, dass es einen Zusammenhang zwischen dem Sortierungskriterium und der untersuchten Eigenschaft gibt. Ein Beispiel: Wir wollen wissen, ob die Hilfedauer mit dem Aufnahmealter zusammenhängt. Dazu bilden wir zwei Altersgruppen (Alter bei der Aufnahme) und vergleichen die mittlere Hilfedauer dieser drei Gruppen. Wenn die Unterschiede signifikant sind, können wir auf einen Zusammenhang schließen.

Tabelle 1: Test eines Zusammenhangs über Mittelwertvergleich

Aufnahmealter	durchschnittliche Hilfedauer
unter 12 Jahre	1,5 Jahre
12 Jahre und älter	1,0 Jahre

N= 5654; der Unterschied ist signifikant $p < 0,001$

Noch einfacher könnte man dieses über eine Korrelation zwischen dem Aufnahmealter und der Verweildauer ausrechnen. Da beide Variablen auf Verhältnisskalen gemessen werden, kann man einen einfachen Korrelationskoeffizienten angeben, der seinerseits wiederum auf Signifikanz geprüft wird. Die Korrelation beträgt -0,15 und ist ebenfalls signifikant ($p < 0,001$). Das negative Vorzeichen besagt, dass mit steigendem Aufnahmealter die Hilfedauer kürzer wird. Korrelationskoeffizienten variieren von -1 bis +1. Der Wert 0,15 ist nicht sehr hoch, sondern deutet auf einen kleinen Zusammenhang hin. Das relativiert den Eindruck aus der Tabelle 1 etwas. Es gibt eben doch viele kurze Hilfen bei jüngeren Kindern und längere Hilfen bei älteren Jugendlichen.

In den Wirkungsberichten stellen wir nur wenige explorative Ergebnisse dar, hauptsächlich im Zusammenhang mit der Effektstärke und der Handlungsrelevanzmatrix, die wir beide unten genauer erklären. Es gibt theoretisch 28321 Zusammenhänge, die untersucht werden könnten. Das alles durchzurechnen, macht keinen Sinn. Vielmehr müssen explorative Analysen auf plausiblen Fragestellungen und Hypothesen beruhen, die wir bis auf einige Ausnahmen dem Leser überlassen wollen. Er hat die Möglichkeit, durch die Gegenüberstellung der zahlreichen Wirkungsberichte auf der CD die wichtigsten Fragen selber zu beantworten. Zwar kann er keine Signifikanzprüfungen durchführen, aber der Blick auf die Fallzahlen wird ihm genug Anhaltspunkte für die

Verlässlichkeit von Unterschieden und Zusammenhängen liefern. Bei speziellen Fragestellungen kann er auf die Rohdaten oder auf die Auswertungsmöglichkeiten des els-Institutes zurückgreifen.

Spannend werden explorative Analysen dann, wenn man komplexere Wirkmodelle überprüft, die mehrere Variable umfassen. Dazu gehören Regressionsanalysen, Clusteranalysen und Faktorenanalysen. Solche multivariaten statistischen Verfahren werden hier nicht dargestellt und erklärt. Multivariate Verfahren brauchen eine große Datenmenge. Wenn zukünftig in Berlin kontinuierlich und flächendeckend Evaluationen durchgeführt werden und die Validität der Daten noch etwas besser wird, können in kurzer Zeit Analysen mit einem hohen Prognosenutzen gemacht werden, die auch im Einzelfall evidenzbasierte Entscheidungshilfen liefern (zum Beispiel durch das Verfahren „decision tree“). Klassische Controllingverfahren in Form von vielen detailreichen Tabellen können dieses nicht leisten. Das vom e/l/s-Institut genutzte Statistik-Programm SPSS (*Statistical Package for the Social Sciences*) bietet viele Möglichkeiten, die von der Firma IBM auch für kommunale Planer angeboten werden. Pro Bezirk würde eine Lizenz für einen geschulten und erfahrenen Statistiker ausreichen. Die wichtigsten Auswertungsroutinen könnte das e/l/s-Institut in Form von Syntax-Skripten zur Verfügung stellen.

Handlungsrelevanzmatrix:

Die einzige multivariate Auswertung, die standardmäßig gemacht wird, ist die Berechnung einer Handlungsrelevanzmatrix (Moisl, 2011). Mit diesem Verfahren kann überprüft werden, bei welchen Dimensionen Handlungsbedarf besteht. Das Verfahren wird in der Organisationsentwicklung zur Fokussierung strategischer Veränderungen (change management) eingesetzt. Bei einer Strategie müssen nämlich zwei Kriterien gleichzeitig bedacht werden: Wo gibt es Abweichungen und als wie wichtig werden diese eingeschätzt.

Bei der WIMES-Handlungsrelevanzmatrix werden zwei Informationen miteinander verknüpft: Wir wirksam ist die Hilfe in einer Dimension? Und: Wie wichtig ist diese Dimension für die Zielerreichung der Hilfe? Wenn Probleme in wichtigen Dimensionen wenig wirksam gelöst werden, besteht ein hoher Handlungsbedarf. Schlechte Ergebnisse in unwichtigen Dimensionen können hingenommen werden, stehen zumindest nicht ganz oben auf der Tagesordnung. Wenn in wichtigen Dimensionen gute Ergebnisse erzielt werden, handelt es sich um eine Stärke der Dienstleistungen.

Wie die Wirksamkeit bezogen auf die einzelnen Dimensionen gemessen wird, ist bereits erklärt worden. Bei der Einschätzung der Wichtigkeit der einzelnen Dimensionen wenden wir einen kleinen statistischen Trick an. Die Alternative bestünde darin, die Bedeutung der Dimensionen im jedem einzelnen Fall abzufragen, was zu einem erheblichen Mehraufwand bei der Datenerhebung führen würde. Der Trick besteht darin, die Gesamtbewertung der Hilfe durch die Fachleute mit den neun Bedarfseinschätzung am Ende zu korrelieren. Wenn es einen hohen Zusammenhang zwischen dem noch bestehenden Bedarf und der Gesamtbewertung der Hilfe gibt, wird die Bedeutung der jeweiligen Dimension als hoch angesetzt. Wenn diese keinen Einfluss auf das Gesamurteil hat, wird sie wohl nicht so wichtig sein. Die theoretische Begründung für dieses Vorgehen liegt darin, dass Urteile in der Regel nicht aus allen zur Verfügung stehenden Informationen gebildet werden, sondern unser Gehirn eine Auswahl trifft und sich auf die wesentlichen Aspekte fokussiert.

Die Handlungsrelevanzmatrix liefert vier Felder unterschiedlicher Bedeutung (Abbildung 7). In diese vier Felder verteilen sich die neun Wirkungsdimensionen. Natürlich lassen sich Maßnahmen nicht direkt aus dieser Analyse ableiten. Dazu bedarf es einer fachlichen Interpretation und Diskussion. Möglicherweise müssen die Gewichte noch einmal überdacht werden und die Aufmerksamkeit mehr auf bestimmte Dimensionen gelenkt werden. Oder es wird deutlich, dass bei der Schlusswertung der Hilfe manche Aspekte zu sehr in den Fokus geraten. In jedem Fall sollten die Punkte nach Möglichkeit von links unten nach rechts oben verlaufen. Ziele, die einem wichtig sind, sollten erfolgreich bearbeitet werden. Was nicht so wichtig ist, muss nicht auf einem hohen Niveau bearbeitet werden.

Das Verfahren ist auch dadurch interessant, dass das Ergebnis nicht bewusst herbeigeführt werden kann. Es setzt sich aus vielen Entscheidungen unterschiedlicher Personen zusammen. Die Berechnung ist so komplex, dass gezielte Veränderungen nur möglich sind, wenn es tatsächlich zu einer Veränderung der Praxis (Performance) und der Haltungen (Relevanz) kommt.

Als Beispiel zeigen wir noch die Handlungsrelevanzmatrix teilstationärer Hilfen in den fünf Bezirken (Abbildung 8) im Vergleich zu den teilstationären Hilfen im übrigen Bundesgebiet (Abbildung 9). Man kann

sehr schön erkennen, dass die Bedeutung der familienbezogenen Zieldimensionen in Berlin deutlich geringer ist. Entsprechend sind auch die familienbezogenen Wirkungen sehr viel schlechter. Welche Schlüsse daraus zu ziehen sind, muss jugendhilfopolitisch entschieden werden.

Abbildung 7: Handlungsrelevanzmatrix und ihre strategische Bedeutung



Abbildung 8: Handlungsrelevanzmatrix bei teilstationären Hilfen im Berliner Projekt

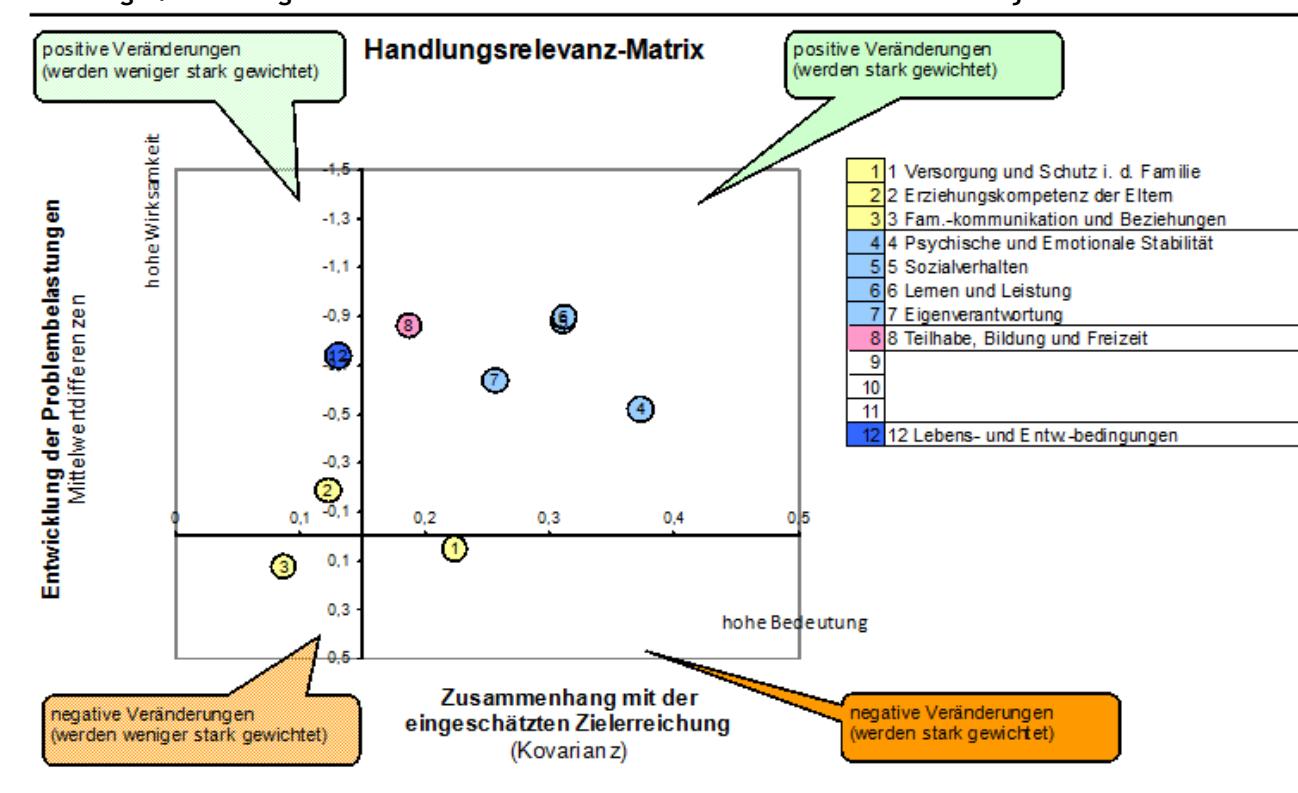


Abbildung 9: Handlungsrelevanzmatrix teilstationärer Hilfen im übrigen Bundesgebiet

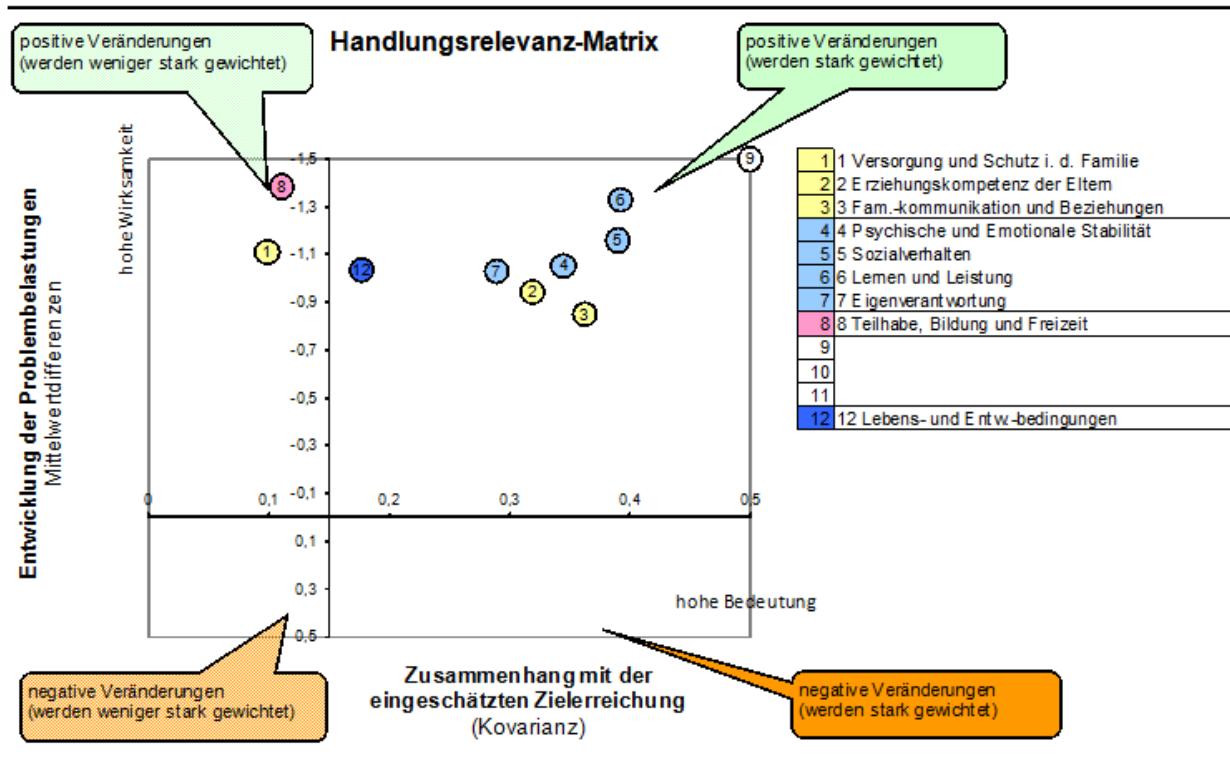


Tabelle 2: Erklärung statistischer Begriffe

Begriff	Erläuterung
Alpha-Fehler	siehe p und Signifikanz
Grundgesamtheit	Gegenstand der Untersuchung oder Evaluation; bei der Evaluation besteht die Grundgesamtheit aus allen gegenwärtigen und zukünftigen Fällen des Systems, das evaluiert werden soll.
Histogramm	Verteilung einer Variable mit kontinuierlichen Messwerten. Die Messwerte werden zu Gruppen zusammengefasst. Die Anzahl der Fälle je Gruppe wird ermittelt und bildet die Höhe einer Säule in einer Häufigkeitsverteilung.
Kategorie	Antwortliste, nach denen Merkmalsausprägungen dokumentiert werden. Nominalskalen brauchen in der Regel einen Katalog von Antwortmöglichkeiten, die alle möglichen Ereignisse abdecken müssen.
Korrelation	Zusammenhang zwischen zwei Messreihen (A und B); r kann Werte zwischen -1 und +1 annehmen; ein negatives Vorzeichen bedeutet: je höher A, desto kleiner B; es gibt unterschiedliche Berechnungsarten, je nach der Art der Verteilungen.
Median	deskriptive Kennzahl einer Verteilung von Messwerten; zentrales Moment; wenn man alle Messwerte aufsteigend sortiert, ist der mittlere Wert der Median; auch Perzentil 50: 50% der gemessenen Werte liegen unter dem Median; der Median ist von der Form der Verteilung unabhängig.
Merkmal	Ausprägung oder Wert einer Variable, die diese in einem konkreten Fall annnehmen kann.
Mittelwert (Mean)	deskriptive Kennzahl einer Verteilung von Messwerten; das arithmetische Mittel errechnet man, indem man alle Messwerte addiert und durch die Anzahl der Messwerte teilt. Mittelwerte setzen symmetrische Verteilungsformen voraus.

Begriff	Erläuterung
N	Anzahl der Datensätze einer untersuchten Einheit (Grundgesamtheit oder Stichprobe)
Nominalskala	Variable mit n Antwortkategorien, denen zur Codierung zwar Zahlen zugeordnet werden können, die Zahlen dürfen aber keinen Rechenoperationen unterzogen werden.
Ordinalskala, Rangskala	Messwerte, bei denen die auf- oder absteigende Anordnung der Zahlen interpretiert werden kann.
p	Abkürzung für die Wahrscheinlichkeit eines Alpha-Fehlers, also die Wahrscheinlichkeit, dass ein Unterschied oder ein Zusammenhang nur auf einen Stichprobenfehler oder einen Messfehler zurückzuführen ist.
relative Häufigkeit	Die Häufigkeit einer Merkmalsausprägung wird durch N geteilt. Die Summe aller relativen Häufigkeiten ist 100%. Bei %-Angaben muss man immer darauf achten, auf welche Stichprobe die Häufigkeiten bezogen werden.
Signifikanz	Überzufälligkeit; zufällige Fallauswahl und Messfehler führen dazu, dass in Stichproben gemessene statistische Kennwerte, Zusammenhänge und Unterschiede nicht die Wirklichkeit widerspiegeln. Mit der Größe der Stichprobe und der Genauigkeit des Mess- und Berechnungsverfahrens steigt die Wahrscheinlichkeit, dass wahre Aussagen über die Grundgesamtheit gemacht werden und die Aussagen prognostischen Wert bekommen. Die statistische Signifikanz wird als Wahrscheinlichkeit angegeben, mit der eine Aussage aufgrund einer Stichprobe nicht den wahren Sachverhalt wiedergibt. $p>0,01$ heißt zum Beispiel, dass eine Aussage mit einer Wahrscheinlichkeit von unter 1 % falsch ist, weil sie auf Zufälle und Messfehler zurückzuführen ist; in den Sozialwissenschaften gibt man sich in der Regel mit $p>0,05$ zufrieden.
Stichprobe	zufällige Auswahl von Fällen aus einer Grundgesamtheit, bei denen ein Kennwert ermittelt wird, oder ein Zusammenhang oder Unterschied untersucht werden soll; je größer die Stichprobe, desto höher die Verlässlichkeit (\Rightarrow Signifikanz); Wenn Stichproben nicht zufällig ausgewählt werden, kann kein Rückchluss auf die Grundgesamtheit gemacht werden.
Streuung = Standardabweichung	Maß für die Unterschiedlichkeit der gefundenen Messwerte; die Einzelwerte "streu" um einen Mittelwert; die Streuung ist gleich der Wurzel aus der \Rightarrow Varianz.
Variable	Eigenschaft eines Gegenstandes oder Prozesses, die variiert, also unterschiedliche Merkmalsausprägungen annimmt. In der explorativen Statistik unterscheidet man „abhängige“ Variable, deren Zusammenhang mit anderen Variablen und Bedingungen man untersuchen möchte, und „unabhängige“ Variable, die möglicherweise einen Einfluss auf die abhängige Variable haben.
Varianz	Maß für die Unterschiedlichkeit der gefundenen Messwerte; alle Abstände zum Mittelwert werden quadriert, addiert und durch die Anzahl der Werte minus 1 geteilt. "Varianz" wird auch umgangssprachlich für die Unterschiedlichkeit von vorgefundenen Merkmalsausprägungen gebraucht.
Verhältnisskala	Abbildung von Merkmalsausprägungen, bei denen das Verhältnis zwischen den Zahlen interpretierbar ist. Klassische Variable, die auf Verhältnisskalen gemessen werden, sind Zeit, Geld, Häufigkeiten. Beispiel: 12 Monate sind doppelt so lang wie 6 Monate.
Verteilung	Liste oder Grafik der Häufigkeiten, mit denen eine bestimmte Merkmalsausprägung einer Variable (Kategorie, Rang, Messwert) auftritt.
Verteilungsparameter	Kenngrößen, mit denen man eine Verteilung beschreiben kann, z.B. Mittelwert, Streuung, Max-Wert, Min-Wert, Schiefe, Form usw.

4.3 Allgemeine Effektstärke aller Erziehungshilfen

Die zentrale Wirksamkeitskennzahl ist die Effektstärke, also die durchschnittliche Differenz zwischen den Bedarfslagen, die zu Hilfen von Erziehung geführt haben (A1) und den erreichten Kompetenzen und Ressourcen am Ende der Hilfe (E). Hierzu werden für aggregierte Gruppen (Hilfearten, Hilfemerkmale, Zielgruppen) Mittelwertvergleiche angestellt. Im Sinne der Wirksamkeit ist es wünschenswert, dass die Differenzen zwischen Anfang und Ende negativ sind, also eine Abnahme von Bedarfen ausweisen. Je negativer die Mittelwertdifferenzen, desto effektiver die Hilfen. Die Unterschiede müssen auf dem 5%-Niveau signifikant sein.

Eine Bestätigung der Signifikanz dürfte bei der hohen Fallzahl nicht schwierig sein. Große Stichproben haben aber den Nachteil, dass auch kleinste Unterschiede signifikant sind, was zu einer gewissen Trivialität der Aussagen führen kann. Wir werden deswegen im Folgenden den Überlegungen von COHEN folgen und nicht nur die statistische Signifikanz, sondern auch die Stärke der Effekte berechnen (Bortz & Döring, 2002). Die Effektstärke nach COHEN's d wird wie folgt berechnet: Der Mittelwertunterschied zwischen Anfang und Ende wird in Beziehung gesetzt zu der Streuung der Anfangswerte. Verändern sich die Mittelwerte durch die Intervention signifikant um mehr als eine halbe Standardabweichung (0,5 sd), so wird eine mittlere Effektstärke angenommen. Beträgt der Abstand zwischen Anfang und Ende mehr als 0,8 sd, so ist die Effektstärke hoch. Statt durch d-Werte werden wir die Effektstärken grafisch darstellen, so dass der Leser auch sehen kann, ob es sehr hohe Effekte gibt oder mitunter auch Negativeffekte, also Verschlimmerungen der Bedarfslagen.

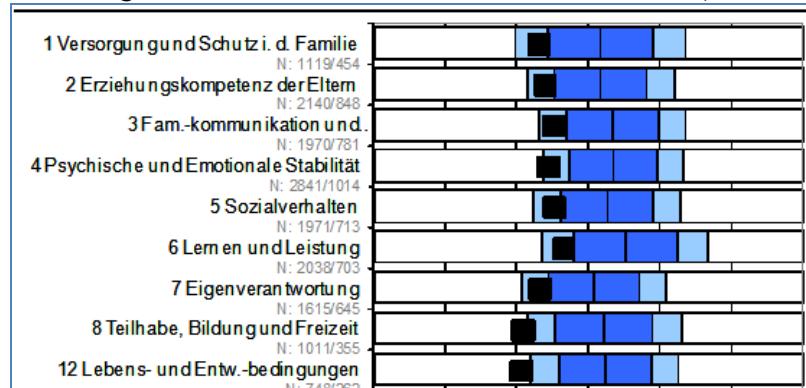
Die Bedarfe wurden auf 9 Dimensionen fachlich eingestuft (siehe Anlage 7). Die Werte können zwischen 1 (kein Bedarf) und 7 (extrem hoher Bedarf) variieren, wobei die Extremwerte im Grunde als Mittelwerte ausfallen. Die mittleren Bedarfe variieren zwischen 2,5 bis 4,5. Besonders in Berlin fällt auf, dass die einzelnen Dimensionen sehr hoch interkorrelieren, dass also ein hoher Bedarf in Dimension A einen hohen Bedarf in der Dimension B mit sich zieht. Dadurch ist es möglich, einen Gesamtbedarf aus allen Dimensionen (= Lebensbereiche von Familien und jungen Menschen) zu berechnen. Über alle Hilfearten, Bezirke und Hilfearten hinweg zeigen sich deutliche Effektstärken (Tabelle 3). Die erwarteten Effektstärken sind damit empirisch bestätigt.

Tabelle 3: Effektstärke bei der allgemeinen Bedarfsreduzierung

Was wird untersucht?	Mittelwert zu Beginn	Mittelwert am Ende	Effektstärke
Bedarfe insgesamt, alle Fälle, alle Hilfen	3,89	3,15	mittelhoch
Bedarfe bei den Richtungszielen, alle Fälle, alle Hilfen	4,30	3,45	hoch

Um aus den Evaluationsergebnissen Erkenntnisse für die Steuerung und Qualitätsentwicklung zu gewinnen, ist dieser allgemeine Befund noch nicht geeignet. Eine Differenzierung nach unterschiedlichen Hilfezielen (hierfür stehen neun Dimensionen zur Verfügung), Hilfearten und Bezirken bringt zusätzliche Erkenntnisse darüber, welche Ziele in welchen Hilfesettings in welchen Organisationen wie erreicht werden. Abbildung 10 zeigt die Effektstärken bei den einzelnen Dimensionen über alle Hilfearten hinweg an. Genutzt werden hier nur die hilfeplanrelevanten Daten, also solche Fälle, bei denen die betreffende Dimension als ein Richtungsziel der Hilfe markiert wurde. Die Effektstärken werden hier grafisch angezeigt: Wenn das schwarze Quadrat (mittlere Bedarfslage am Ende) im hellblauen Teil des Balkens steht, zeigt das eine mittlere Effektstärke an. Die Lage weiter links außerhalb des hellblauen Balkens verweist auf große Effektstärken.

Abbildung 10 Effekte in den einzelnen Zieldimensionen; alle Hilfen, alle Daten



Am höchsten sind die Effekte bei der Integration und der Lebenssituation der jungen Menschen sowie bei deren psychischer Stabilität. Die Zahlen unter den Dimensionsbezeichnungen geben an, in wie vielen Fällen die betreffende Dimension als Richtungsziel geplant worden ist. Am häufigsten geht es um die psychische Verfassung der jungen Menschen, ihre Leistungs- und Lernfähigkeiten und um die Erziehungskompetenz der Eltern. Seltener geht es dagegen um die Sicherung der Kinderrechte und um die Themen Integration und Inklusion.

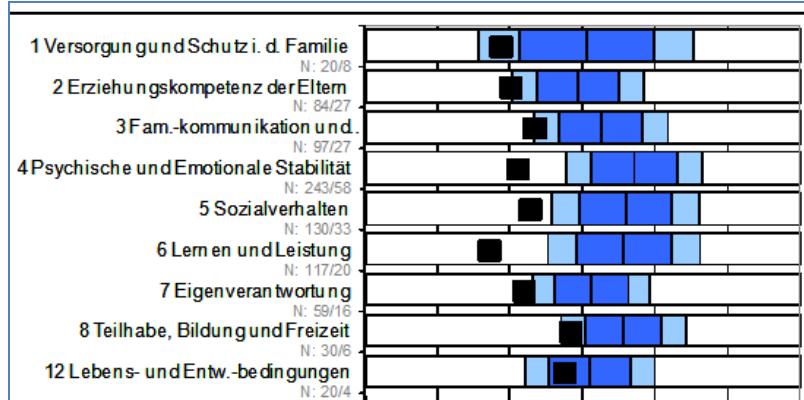
Die Befunde variieren bei den einzelnen Hilfearten stark, sowohl bei den Ausgangsbedarfen als auch bei den erzielten Effekten. Die Gesamteffektstärken, differenziert nach den Berliner Leistungstypen, sind in der Tabelle 4 dargestellt. Auch hier werden nur die Zieldimensionen der Hilfen dargestellt.

Tabelle 4: Ausgangsbedarfe und Effektstärke nach Leistungstypen; alle 5 Bezirke; Richtungsziele

Leistungstyp	Bedarf A1	sd A1	mittlerer Effekt	Cohen's d	Effektstärke
amb PsychTher	4,42	1,20	-1,27	1,1	hoch
integr LernTher	4,42	1,32	-1,68	1,3	hoch
amb FamTher	4,06	,99	-1,02	1,0	hoch
Soz GrA	3,96	1,06	-,72	0,7	mittel
ErzBei	4,06	1,16	-,82	0,7	mittel
SPFH	4,14	1,18	-,89	0,7	mittel
amb ISE	4,69	1,24	-,61	0,5	klein bis mittel
teilstationär	4,54	1,01	-,84	0,8	mittel bis hoch
stat (Gruppe)	4,79	1,14	-,87	0,8	mittel bis hoch
stat (WG)	4,48	1,23	-,86	0,7	mittel
stat fam	4,70	1,26	-,97	0,8	mittel bis hoch
stat Ind BEW	4,14	1,24	-,66	0,5	klein bis mittel
stat ISE	5,05	1,07	-1,13	1,1	hoch
Sonstige stationär	4,78	1,25	-1,03	0,8	mittel bis hoch
Sonstige ambulant	4,10	1,24	-1,20	1,0	hoch
Sonstige teilstationär	4,26	1,10	-,18	0,2	klein

Als Beispiel eines Hilfesettings mit großen Effekten sei hier die ambulante Psychotherapie dargestellt (Abbildung 11). Hauptsächlich geht es hier um die psychische Stabilisierung des jungen Menschen, die dann auch große Veränderungen in den Bereichen Schule und Sozialverhalten bewirkt. Familie und Umfeld sind weniger im Fokus.

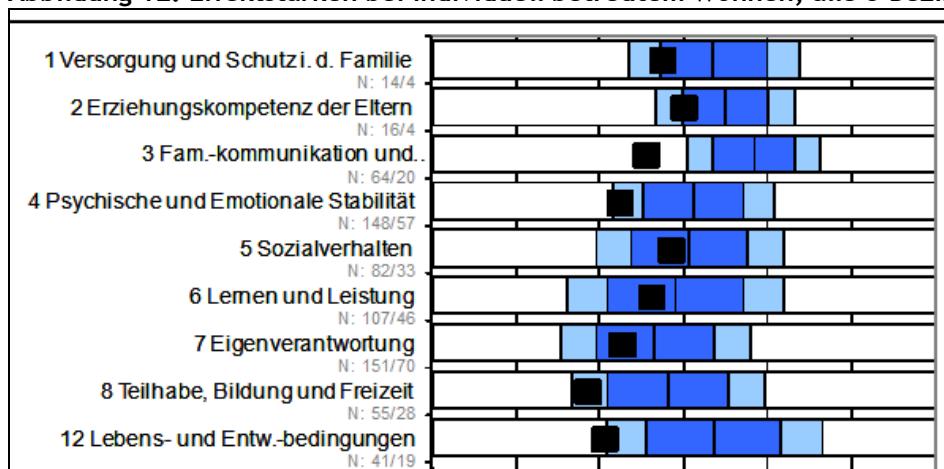
Abbildung 11: Effekte bei den Richtungszielen ambulanter Psychotherapie



Weniger groß sind die Effekte beim individuell betreuten Wohnen (BEW; Leistungstyp 12)(Abbildung 12). Dieses liegt hauptsächlich daran, dass die Entwicklung der jungen Menschen keine großen Fortschritte mehr macht. Zwar lassen sich bezogen auf die Herkunftsfamilie wohl noch einige Klärungen erreichen und insgesamt scheint es den jungen Menschen besser zu gehen als vorher. Aber gerade in den Bereichen Sozialverhalten, Arbeiten und Lernen und in der Selbstständigkeit sind die Effekte eher klein. Diese Richtungsziele werden aber am häufigsten vereinbart und bei der Bewertung der Hilfe durch die Fachleute besonders stark herangezogen.

Bei einer Interpretation und Bewertung solcher Befunde sollte man vorsichtig sein. Möglicherweise ist der Einfluss auf die Persönlichkeitsentwicklung nach einer stationären Hilfe eher gering, weil jetzt das Interesse der jungen Menschen wohl mehr darauf ausgerichtet ist, endlich einmal in Ruhe gelassen zu werden. Möglicherweise reichen die Ziele aus, jetzt eine Reintegration in das „normale“ Leben zu erreichen und den Rückfall in alte Verhaltensweisen zu verhindern. Von normativ erzieherischen Absichten in Richtung Leistungsfähigkeit und „Wohlanständigkeit“ sollte man in dieser Entwicklungsphase in pädagogischer Selbstbescheidenheit wohl Abstand nehmen. Es hat übrigens keinen Einfluss, ob diese Richtungsziele vereinbart wurden oder nicht.

Abbildung 12: Effektstärken bei individuell betreutem Wohnen; alle 5 Bezirke; Richtungsziele



Zusammenfassung und Hinweise:

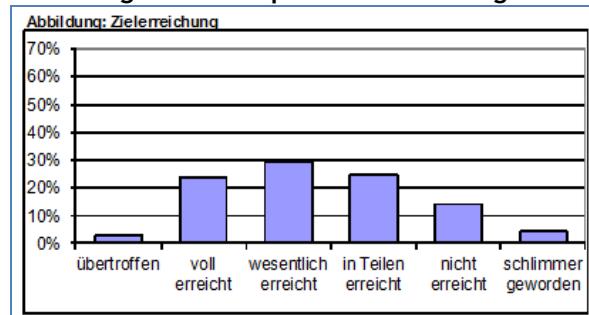
Die Effektstärke Berliner Hilfen ist insgesamt befriedigend bis gut. Eine generelle Problemanzeige ergibt sich daraus nicht. Am besten schneiden die therapeutischen Hilfen ab. Relativ gering sind die Effektstärken bei den teilstationären und stationären Hilfen, die allerdings auch mit einem höheren Bedarf starten. Bei der Interpretation der Effektstärken muss jeweils sorgfältig beachten werden, ob möglicherweise die Eingangsbewertungen falsch lagen, ob realistische Ziele gewählt wurden oder ob die Effektstärken dadurch gesenkt

werden, dass der Anteil abgebrochener Hilfen hoch ist, wie wir für die teilstationären und stationären Hilfen zeigen werden.

4.4 Retrospektive Einschätzung der Zielerreichung der Hilfe

Diese Kennzahl ist fehleranfällig, weil am Ende oft nicht mehr richtig erinnert wird, was genau die Ziele und die Ausgangssituation waren. Oft dominieren die letzten Eindrücke im Hilfeprozess. Außerdem ändern sich Ziele im Verlauf. Diese Kennzahl wird aussagekräftiger, wenn die Zielerreichung anhand der konkreten Handlungsziele in den Hilfeplangesprächen aus verschiedenen Perspektiven beurteilt wird. In der Projektphase wurde der Verlauf (z. B. mittels WIMES plus) nicht eingeschätzt. Dennoch kann die allgemeine Beurteilung als Hinweise auf die Ergebnisqualität der Hilfe genutzt werden. Abbildung 13 zeigt die durchschnittliche Einschätzung aller Hilfen.

Abbildung 13: Retrospektive Bewertung der Zielerreichung durch die Hilfe; alle Hilfen, alle 5 Bezirke

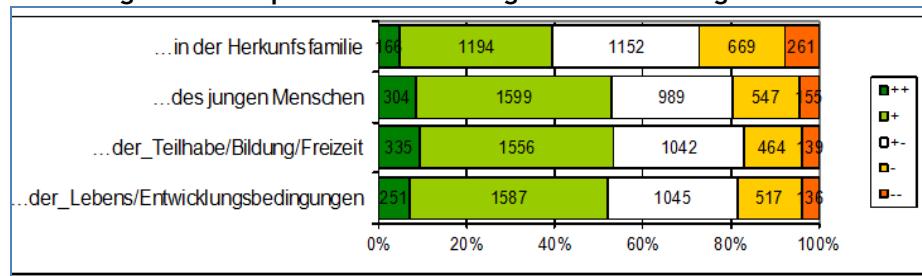


Faustregel zur Bewertung der Ergebnisse: Solange die relative Häufigkeit der misslungenen Hilfen unter 20 % liegt, besteht kein dringender Handlungsbedarf (Pareto-Prinzip). Die Summe der eindeutig gelungenen Hilfen sollte mindestens bei ungefähr 60 % liegen. In Berlin gibt es demnach keine Hinweise auf gravierende Mängel.

4.5 Retrospektive Beurteilung der Entwicklung während der Hilfe

Hier gilt das Gleiche wie oben: Die gelben und roten Einschätzungen (Verschlechterungen) sollten nicht über 20 % liegen. Die grünen Einschätzungen mindestens bei 60 %. Auch hier die Einschätzung über alle Hilfen (Abbildung 14):

Abbildung 14: Retrospektive Beurteilung der Entwicklungen während der Hilfe (alle Hilfen)



4.6 Der Anteil der nicht planmäßig beendeten Hilfen (Abbruchquote)

Am Ende der Hilfen wird von den Fachleuten beurteilt, ob die Hilfe wie geplant und im Konsens beendet wurde, oder ob die Hilfe vorzeitig im Dissens abgebrochen wurde. Die Beendigungsart korreliert sehr hoch mit dem Erfolg der Hilfe. Hilfen, die wie geplant durchgeführt wurden, sind in der Regel erfolgreich, während ein Abbruch fast immer mit einem Misserfolg gleichzusetzen ist.

Einige Fallverantwortliche hatten Probleme damit, die Definition der Beendigungsart richtig anzuwenden⁴. Wenn im Konsens ein anderes Hilfesetting als geeigneter angesehen wird, ist dieses auch als geplante Beendigung einzuschätzen. Das ist nicht von allen so gehandhabt worden. Es kann also sein, dass die Abbruchquote dadurch etwas zu hoch ausfällt. Andererseits ist eine nötige Verlegung auch nicht gerade ein erwünschtes Ergebnis. Insofern kann die Abbruchquote, so wie sie im Berliner Projekt ermittelt wurde, durchaus als Wirkungsindikator verwendet werden. Tabelle 5 zeigt die Abbruchquoten einiger ausgewählter Hilfearten.

Tabelle 5: Abbruchquote im Berliner Projekt (nicht gewichtete Stichprobe)

Hilfeart	Abbruchquote	... im Vergleich zum bundesweiten Benchmark
alle Hilfen zusammen	30,5 %	etwas höher
alle stationären Hilfen	42,9 %	höher
teilstationäre Hilfen in Tagesgruppen	40,4 %	viel höher
ambulante therapeutische Hilfen	18,8 %	(keine Vergleichszahlen)
ambulante sozialpädagogische Hilfen	26,5 %	ungefähr gleich

Wenn man die Effektivität der abgebrochenen Hilfen mit der geplant durchgeföhrter Hilfen vergleicht (Tabelle 6), wird deutlich, dass der stärkste Hebel zur Verbesserung der Ergebnisqualität darin liegt, Hilfen so zu planen, zu steuern und durchzuführen, dass die Stabilität des Hilfeverlaufs möglichst hoch ist. Wahrscheinlich muss Jugendhilfe mit einer Abbruchbruchquote von 20 % und bei besonders schwieriger Klientel auch von 30 % leben - wir kennen den im Bestfall zu erreichenden Wert nicht.

Tabelle 6: Effektivität der Hilfesettings je nach Art der Beendigung

Stichprobe	abgebrochen		wie geplant beendet	
ambulant ther.	4,26 → 3,65	mittlerer Effekt	4,26 → 2,72	hoher Effekt
ambulant sozpd.	4,11 → 3,99	kein Effekt	4,11 → 2,86	hoher Effekt
teilstationär	4,54 → 4,37	kein Effekt	4,54 → 3,23	hoher Effekt
stationär	4,64 → 4,36	kleiner Effekt	4,64 → 3,05	hoher Effekt

Zusammenfassung und Hinweise:

Die retrospektiven Kennzahlen kommen im Wesentlichen zu den gleichen Ergebnissen wie die prospektiven Kennzahlen. Sie können also durchaus zur Wirkungsevaluation herangezogen werden. Insbesondere ist die Frage nach der Art der Beendigung aufschlussreich. In Berlin sind die Abbruchquoten bei den teilstationären und stationären Hilfen auffällig hoch. Dieses führt zu geringeren durchschnittlichen Effektstärken, mehr Folgehilfen und höheren Fallzahlen insgesamt. Die Abbruchquote ist die steuerungsrelevanteste Kennzahl und wird deswegen in der Folge noch weiter differenziert und analysiert.

4.7 Folgehilfen

Am Ende einer Hilfe wird danach gefragt, ob und ggf. welche Hilfe zur Erziehung in der Folge geplant ist. Daraus lässt sich pro Hilfeart oder Bezirk die relative Anzahl der Folgehilfen errechnen. Es muss allerdings beachtet werden, dass dieser Indikator einer Verzerrung unterliegt: Bei abgebrochenen Hilfen kommt es nicht zu einer unmittelbaren Folgehilfe, weil es für eine solche Planung keine Gelegenheit gibt. Es ist wahrscheinlich, dass im Falle eines Abbruchs später wieder eine Hilfe nötig wird, den genauen Anteil kann man

⁴ Aus diesem Grund wird in der WIMES Web-Portal Version 2013.1 die Frage nach der Beendigung der Hilfen differenzierter formuliert:

1. Beendigung in Übereinstimmung mit der bisherigen Hilfeplanung (Eingangsplan und Anpassungen im Verlauf),
2. Abbruch der laufenden Hilfe durch faktische Beendigung im Dissens bei weiter bestehendem Bedarf,
3. Beendigung durch externe Umstände (Umzug, Krankheit, Sorgerechtsänderung, Haft ...),
4. Bedarfsgerechte Veränderung der Hilfeart und des Hilfesettings (Folgeintervention siehe unten).
- 5.

aber erst ermitteln, wenn die spätere Entwicklung nachverfolgt wird, also in einer Längsschnittuntersuchung mit der Überprüfung der Nachhaltigkeit.

Außerdem ist die Bewertung der Kennzahl nicht ganz einfach. Zwar ist es wünschenswert, wenn mittels einer Hilfe die Bedarfe soweit abgebaut werden konnten, dass keine Folgehilfe nötig wurde. Dennoch gibt es viele Verläufe, bei denen eine Folgehilfe als Erfolg eingestuft werden müsste. Zum Beispiel ist es fachlich positiv zu bewerten, wenn im Anschluss einer Heimerziehung ein Betreutes Wohnen zur Verselbständigung durchgeführt wird, oder wenn durch eine ambulante oder teilstationäre Hilfe so viel Einsicht oder Vertrauen bei den Eltern erzeugt werden konnte, dass eine notwendige stationäre Hilfe durchgeführt werden kann. Umgekehrt ist es fachlich nicht wünschenswert, wenn Klienten in einer Misserfolgskarriere eine Hilfe an die nächste reihen oder wenn eine präventiv angelegte ambulante Hilfe letztlich eine stationäre Unterbringung doch nicht hat verhindern können.

Für eine aussagekräftige Kennzahl bezogen auf die Folgehilfen wäre es nötig, im Rahmen des zukünftigen Fachverfahrens nachzuverfolgen⁵, welche Hilfen zu einem späteren Zeitpunkt nötig sind. Diese Daten müssen aber mit Informationen verknüpft werden, ob dieses im Sinne des Kindeswohls eher positiv oder eher negativ zu bewerten ist. Außerdem wäre es interessant, auch andere Folgehilfen außerhalb der Jugendhilfe zu dokumentieren⁶. Beschränken wir uns hier zunächst auf die deskriptive Kennzahl (Abbildung 15 bis Abbildung 17).

Abbildung 15: Nachfolgende Hilfen bei allen beendeten stationären Hilfen

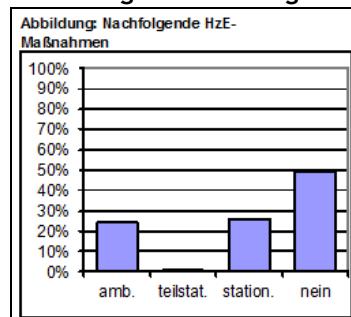
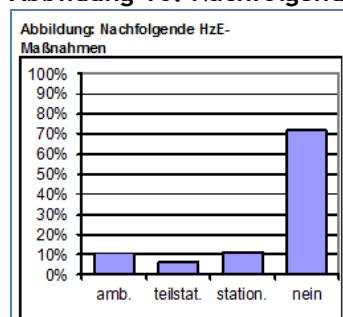
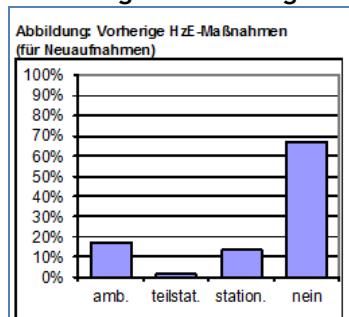


Abbildung 16: Nachfolgende Hilfen bei allen beendeten teilstationären Hilfen



⁵ Eine entsprechende Erhebung mittels WIMES ist von Datenschützern mit dem Hinweis auf eine verbotene Vorratsdatenspeicherung unterbunden worden.

⁶ In der nächsten WIMES-Version 2013.1 wird der Katalog der Folgehilfen erweitert: stationäre HzE, teilstationäre HzE, ambulante HzE, sozialräumliche Lösung, schulische oder berufsvorbereitende Hilfen, Kinder- und Jugendpsychiatrie / Therapie, Eingliederungshilfe (SGB XII), Haft / Strafvollzug, keine

Abbildung 17: Nachfolgende Hilfen bei allen beendeten ambulanten Hilfen**Zusammenfassung und Hinweise:**

Die nachfolgend geplanten Hilfen sind im Vergleich zu der Benchmark-Stichprobe nicht auffällig. Hierbei muss allerdings beachtet werden, dass bei den abgebrochenen Hilfen logischer Weise keine weiteren Planungen erfolgen, dass es aber wahrscheinlich zu späteren Zeitpunkten erneut zu Erziehungshilfen kommt. Deshalb ist es ratsam, diese Kennzahl im Fachverfahren zu dokumentieren und in Kennzahlen für das Fachcontrolling darzustellen. Aufgrund der Anonymisierung der Fälle können langfristige Folgehilfen mit dem Evaluationsverfahren nicht festgestellt werden.

4.8 Weitere mögliche Wirkungskennzahlen, die im Pilotprojekt nicht erhoben worden sind*Zielerreichung der Zielperspektive*

Zur Überprüfung der Zielerreichung ist es nötig, die Zielperspektive, wie sie am Anfang festgelegt wurde, in Beziehung zu setzen zu dem tatsächlichen Verlauf und der Zielerreichung, wie sie sich am Ende darstellen. Hierzu fehlt in dem dokumentierten Zeitraum mindestens ein Jahr. Außerdem empfiehlt es sich, eventuelle Korrekturen bei der Zielperspektive im Hilfeverlauf zu erfassen. Auch dazu gibt der Berliner Datensatz zurzeit noch keine Informationen.

Wir können hier lediglich die Empfehlung aussprechen, diese Indikatoren zu erheben, weil sie den Vorteil haben, auf gut zu beobachtenden Daten zu beruhen.

- Bei der Zielperspektive „Verbleib in der Familie“ wird der relative Anteil der Hilfen berechnet, bei denen am Ende oder später keine Fremdunterbringung erfolgt ist.
- Bei der Zielperspektive „Rückkehr in die Familie“ wird der relative Anteil der Hilfen berechnet, bei denen innerhalb von 1,5 Jahren / 2 Jahren eine positiv zu bewertende Rückführung gelungen ist.
- Bei der Zielperspektive „Beheimatung“ wird der relative Anteil der Hilfen berechnet, bei denen der junge Mensch nach 2 Jahren noch stabil in der Hilfe integriert ist.
- Bei der Zielperspektive „Verselbständigung“ wird der Anteil der Hilfen berechnet, an deren Ende der junge Mensch keine Folgehilfe mehr braucht und nicht in einer Institution lebt.

Durchschnittliche Zielerreichungsgrade der Handlungsziele im Hilfeverlauf

Hierzu muss der Verlauf dokumentiert werden und in den Hilfeplangesprächen die Zielerreichung in Bezug auf die letzte Zielvereinbarung eingestuft werden. Voraussetzung sind klare operationalisierte Zielvereinbarungen mit den Klienten und die verlässliche Überprüfung der Zielerreichung in definierten Zeiträumen, mindestens aber im nächsten Hilfeplangespräch.

Da es in dem Projekt nicht mehr zu einer Integration der Wirkungsevaluation in das Hilfeplanverfahren gekommen ist, kann dazu keine Aussage gemacht werden. Dieser Index wird zurzeit im Zusammenhang mit der Entwicklung der Version WIMES plus und der Integration in das Berliner Hilfeplanverfahren vorbereitet.

Konstanz der Hilfe bezüglich der Richtungsziele

Die Nachhaltigkeit und Ernsthaftigkeit der Hilfeplanung sowie die Stabilität der Zielvereinbarungen und der Hilfe insgesamt sind starke Wirkungsfaktoren. Insofern ist dieser Index nicht nur ein Wirkungsindikator, sondern auch ein Frühindikator zur Sicherung der Prozessqualität. Berechnet wird, wie oft Richtungsziele nicht weiter verfolgt werden, weil sie a) vergessen wurden, b) nicht mehr relevant sind, c) als unerreichbar und nicht konsensfähig abgewählt wurden. Diese Indikatoren setzen ebenfalls die Integration wirkungsorientierter Dokumentation in das Hilfeplanverfahren voraus.

Nutzwert aus der Sicht der Hilfeempfänger

Es sollte ein Standard der Hilfeplanung sein, am Ende der Hilfe Klienten nach ihrer Zufriedenheit zu fragen. Die Befragung sollte mit einem einfachen kurzen Fragebogen erfolgen, der die Sprache der Klienten spricht, die Perspektiven der Eltern und der Kinder und Jugendlichen einzeln erfasst und anonym ausgefüllt und ausgewertet wird. Entsprechende Bögen und Auswertungsroutinen und Benchmarks bietet das WIMES-Verfahren an. Allerdings müssen die Fallverantwortlichen die Befragung organisatorisch sicherstellen und die Klienten persönlich motivieren.

Für das Berliner Pilotprojekt wurde eine Klientenbefragung zunächst ausgeschlossen.

Prozesskennzahlen (Indikatoren für die Prozessqualität in der Fallsteuerung)

- Einhalten der Prozessschritte und der Terminvorgaben
- Überprüfung der sozialpädagogischen Diagnose nach 8 Wochen
- Durchführung eines Abschlussgesprächs mit gemeinsamer Ergebnisbewertung und gegenseitiger Rückmeldung
- Durchführung einer Abschlussbewertung in WIMES

4.9 Leistungstypen und Hilfearten differenzierte Wirkungskennzahlen

Bisher wurde relativ ausführlich über die allgemeinen Evaluationsbefunde in den fünf Projektbezirken berichtet, ungeachtet der Berliner Leistungstypen und Hilfearten. In der Folge werden die Evaluationsergebnisse differenzierter betrachtet. Allerdings geben wir hier nur einen Überblick über die zentralen Befunde. Wer sich im Einzelnen über die Auswertungen informieren will, wird auf die Wirkungsberichte auf der beigelegten CD-ROM verwiesen. Der hier gegebene Überblick soll gewährleisten, dass keine wichtigen Einzelbefunde übersehen werden.

Erläuterungen zu den Kennzahlen:

- Anteil Folgehilfen HzE: Zusammenfassung ambulanter, teilstationäre und stationärer Folgehilfen.
- Anteil positiver Zielerreichung: Das bezieht sich auf die (retrospektive) Frage nach der Zielerreichung der Hilfe. Anteil der Fälle, bei denen „übertragen“, „voll erreicht“ oder „wesentlich erreicht“ angekreuzt wurde.
- Anteil misslungener Hilfen: Das bezieht sich auf die (retrospektive) Frage nach der Zielerreichung der Hilfe. Anteil der Fälle, bei denen „nicht erreicht“ oder „schlimmer geworden“ angekreuzt wurde.
- Positive Entwicklungen in Familie: Das bezieht sich auf die (retrospektive) Beurteilung der Entwicklung der Familie während der Hilfe. Anteil der Fälle, bei denen „++“ oder „+“ und die Hälfte der Fälle bei denen „+-“ angegeben wurde.
- Positive Entwicklungen beim jungen Menschen: Das bezieht sich auf die (retrospektive) Beurteilung der Entwicklung des jungen Menschen während der Hilfe. Anteil der Fälle, bei denen „++“ oder „+“ und die Hälfte der Fälle bei denen „+-“ angegeben wurde.
- Negativentwicklungen in Familie: Das bezieht sich auf die (retrospektive) Beurteilung der Entwicklung der Familie während der Hilfe. Anteil der Fälle, bei denen „--“ oder „-“ angegeben wurde.
- Negativentwicklung beim jungen Menschen: Das bezieht sich auf die (retrospektive) Beurteilung der Entwicklung des jungen Menschen während der Hilfe. Anteil der Fälle, bei denen „--“ oder „-“ angegeben wurde.

Alle Tabellen sind gleich aufgebaut, um das Lesen zu erleichtern. Unter den Tabellen wird in Zusammenfassung und Hinweisen deutlich gemacht, wo es Auffälligkeiten und Abweichungen gibt, die möglicherweise steuerungsrelevant sind. Gegebenenfalls werden auch Hypothesen zu den Ursachen genannt und Hinweise zu sich daraus ergebenden Fragestellungen gegeben. Hinweise zu fachlichen Strategien ergeben sich meistens nicht unmittelbar aus den Kennzahlen, sollen aber zu fachlichen Auseinandersetzungen über mögliche Lösungsansätze anregen.

Innerhalb der Tabellen wird grundsätzlich unterschieden zwischen der Auswertung aller Hilfen dieser Hilfeform und einer weiteren Filterung nach der Beendigungsart. In der Unter-Stichprobe der abgebrochenen Hilfen wird deutlich, wie stark Abbrüche die Gesamtergebnisse negativ beeinflussen. Umgekehrt kann man bei den wie geplant beendeten, stabil durchgeföhrten Hilfen sehen, was unter günstigen Bedingungen erreicht werden kann.

4.9.1 Alle ambulanten therapeutischen Hilfen

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	11
Geschlecht m:w	60:40
durchschnittliche Hilfedauer (Monate)	19,6
Abbruchquote	18,8%
Anteil Folgehilfen HzE	51,5%
Anteil positiver Zielerreichung	75,3%
Anteil misslungener Hilfen	8,4%
Positive Entwicklungen in Familie	75%
Positive Entwicklungen beim jungen Menschen	78%
Negativentwicklungen in Familie	16%
Negativentwicklung beim jungen Menschen	12%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	11,2
Anteil misslungener Hilfen	38%
Effektstärke bei Familienzielen	keine
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	mittel
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	24
Anteil positiver Zielerreichung	90%
Effektstärke bei Familienzielen	hoch
Effektstärke bei Zielen junger Menschen	sehr hoch
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

Die Wirksamkeit der ambulanten Hilfen zeigt sich in allen Kennzahlen als sehr hoch. Ambulante Hilfen laufen in einem hohen Maße stabil mit einem hohen Anteil positiver Entwicklungen in allen Lebensbereichen. Auch bei abgebrochenen Hilfen sind immer noch viele Teilerfolge zu erkennen. Erklärungsbedürftig ist bei diesem hohen Wirkungsgrad aber der hohe Anteil der nachfolgenden Hilfen zur Erziehung. Eine zu prüfende Hypothese ist, dass möglicherweise in dem therapeutischen Prozess und auf der Basis eines hohen persönlichen Vertrauens Bedarfe zur Sprache kommen, die vorher nicht sichtbar waren.

4.9.2 Im Einzelnen: ambulante Psychotherapie (Leistungstyp 1)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	11,3
Geschlecht m:w	67:33
durchschnittliche Hilfedauer (Monate)	18,4
Abbruchquote	20,5%
Anteil Folgehilfen HzE	51,5%
Anteil positiver Zielerreichung	73,4%
Anteil misslungener Hilfen	8,1%
Positive Entwicklungen in Familie	68%
Positive Entwicklungen beim jungen Menschen	75%
Negativentwicklungen in Familie	18%
Negativentwicklung beim jungen Menschen	14%
Effektstärke bei Familienzielen	hoch
Effektstärke bei Zielen junger Menschen	sehr hoch
Effektstärke bei Integrationszielen	sehr hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	9,6
Anteil misslungener Hilfen	31,3%
Effektstärke bei Familienzielen	gering bis mittel
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	mittel
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	24
Anteil positiver Zielerreichung	87,5%
Effektstärke bei Familienzielen	hoch
Effektstärke bei Zielen junger Menschen	sehr hoch
Effektstärke bei Integrationszielen	sehr hoch

Zusammenfassung und Hinweise:

Bei den ambulanten Psychotherapien sind die Effekte sehr hoch. Auch hier stellt sich allerdings die Frage, wieso dennoch in so vielen Fällen eine weitere ambulante Erziehungshilfe folgt. Des Weiteren sollte untersucht werden, warum die Abbrüche relativ spät erfolgen. Möglicherweise sollten Beendigung proaktiver angegangen werden, wenn zum Beispiel seitens der Klienten keine weitere Bereitschaft mehr besteht oder die erreichten Teilziele ausreichen.

4.9.3 Im Einzelnen: integrative Lerntherapie (Leistungstyp 2)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	10,6
Geschlecht m:w	56:44
durchschnittliche Hilfedauer (Monate)	24
Abbruchquote	14,9%
Anteil Folgehilfen HzE	64%
Anteil positiver Zielerreichung	81,5%
Anteil misslungener Hilfen	5,5%
Positive Entwicklungen in Familie	80%
Positive Entwicklungen beim jungen Menschen	80%
Negativentwicklungen in Familie	8%
Negativentwicklung beim jungen Menschen	8%
Effektstärke bei Familienzielen	/
Effektstärke bei Zielen junger Menschen	sehr hoch
Effektstärke bei Integrationszielen	hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	14,2
Anteil misslungener Hilfen	31,4%
Effektstärke bei Familienzielen	/
Effektstärke bei Zielen junger Menschen	gering bis mittel
Effektstärke bei Integrationszielen	hoch
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	24
Anteil positiver Zielerreichung	93,6%
Effektstärke bei Familienzielen	/
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	mittel

Zusammenfassung und Hinweise:

Integrative Lerntherapien zeigen eine sehr hohe Wirksamkeit. Es ist die stabilste Hilfeform von allen Leistungstypen. Die Ziele sind eindeutig auf die schulischen Probleme fokussiert. Hier zeigen sich extrem hohe Effektstärken. Zu diskutieren wäre auch hier, warum in einem so hohen Anteil von Fällen nachfolgende Erziehungshilfen nötig sind.

4.9.4 Im Einzelnen: ambulante Familientherapie (Leistungstyp 3)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	12,2
Geschlecht m:w	54:46
durchschnittliche Hilfedauer (Monate)	12,2
Abbruchquote	23,7%
Anteil Folgehilfen HzE	22,6%
Anteil positiver Zielerreichung	65,3%
Anteil misslungener Hilfen	19,4%
Positive Entwicklungen in Familie	55%
Positive Entwicklungen beim jungen Menschen	59%
Negativentwicklungen in Familie	30%
Negativentwicklung beim jungen Menschen	26%
Effektstärke bei Familienzielen	sehr hoch
Effektstärke bei Zielen junger Menschen	mittel bis hoch
Effektstärke bei Integrationszielen	sehr hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	6
Anteil misslungener Hilfen	65,2%
Effektstärke bei Familienzielen	negativ
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	/
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	12
Anteil positiver Zielerreichung	85,3%
Effektstärke bei Familienzielen	sehr hoch
Effektstärke bei Zielen junger Menschen	mittel bis hoch
Effektstärke bei Integrationszielen	sehr hoch

Zusammenfassung und Hinweise:

Ambulante Familientherapien haben eine sehr hohe Effektivität, wenn sie stabil durchlaufen. Im Gegensatz zu den anderen ambulanten Hilfen gibt es seltener Folgehilfen. Abbruch ist eindeutig mit Misserfolg gleichzusetzen. Positiv formuliert: Hilfen, die keine Aussicht auf Erfolg haben, werden rechtzeitig beendet. Die Abbruchquote ist zwar die höchste bei den ambulanten Therapien, liegt aber im Vergleich zu anderen Hilfen zur Erziehung moderat niedrig.

4.9.5 Ambulante sozialpädagogische Hilfen

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	10,6
Geschlecht m:w	60:40
durchschnittliche Hilfedauer (Monate)	11
Abbruchquote	26,5%
Anteil Folgehilfen HzE	25%
Anteil positiver Zielerreichung	56,9%
Anteil misslungener Hilfen	17,5%
Positive Entwicklungen in Familie	60%
Positive Entwicklungen beim jungen Menschen	73%
Negativentwicklungen in Familie	24%
Negativentwicklung beim jungen Menschen	18%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	mittel
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	6,9
Anteil misslungener Hilfen	50%
Effektstärke bei Familienzielen	keine
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	12
Anteil positiver Zielerreichung	71,5%
Effektstärke bei Familienzielen	hoch
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

Bei den ambulanten sozialpädagogischen Hilfen (in Abgrenzung zu den therapeutischen) ist die Wirksamkeit hoch, wenn die Hilfen wie geplant stabil laufen. Da mehr als ein Viertel der Hilfen vorzeitig abgebrochen wird, reduziert sich die Effektivität auf ein mittleres Ausmaß. Da die retrospektiven Einschätzungen positiver als die prospektiven Kennzahlen ausfallen, müsste untersucht werden, ob möglicherweise anfängliche Bedarfe unterschätzt werden. Klären lässt sich dieser Widerspruch am besten, wenn die Zielplanungen genauer in den Blick genommen werden, und wenn untersucht wird, welches die Ursachen für Abbrüche sind. Zwar stellen diese in ihrem Anteil kein so gravierendes Problem dar wie bei den teilstationären und stationären Hilfen, da die ambulanten Hilfen aber einen großen Anteil aller Hilfen ausmachen, würden sich eine Stabilisierung (sprich: Senkung der Abbruchquote) und eine höhere Zielgenauigkeit positiv auswirken.

4.9.6 Im Einzelnen: Soziale Gruppenarbeit (Leistungstyp 4)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	10,2
Geschlecht m:w	66:34
durchschnittliche Hilfedauer (Monate)	12,9
Abbruchquote	38,7%
Anteil Folgehilfen HzE	23,4%
Anteil positiver Zielerreichung	53,3%
Anteil misslungener Hilfen	20,3%
Positive Entwicklungen in Familie	62%
Positive Entwicklungen beim jungen Menschen	70%
Negativentwicklungen in Familie	21%
Negativentwicklung beim jungen Menschen	16%
Effektstärke bei Familienzielen	gering bis mittel
Effektstärke bei Zielen junger Menschen	gering bis mittel
Effektstärke bei Integrationszielen	mittel bis hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	11,8
Anteil misslungener Hilfen	41,9%
Effektstärke bei Familienzielen	negativ
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	hoch
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	16
Anteil positiver Zielerreichung	73,6%
Effektstärke bei Familienzielen	mittel bis hoch
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	sehr hoch

Zusammenfassung und Hinweise:

Für eine ambulante Hilfe ist die Abbruchquote ziemlich hoch. Das mindert die gute Wirksamkeit erheblich, insbesondere auch deswegen, weil die Abbrüche relativ spät erfolgen. Die Fragestellung, die sich daraus ergibt: Wie kann erreicht werden, dass junge Menschen bis zum Schluss in der Hilfe bleiben. Die Negativentwicklungen in den Familien bei den abgebrochenen Hilfen legen den Verdacht nahe, dass die Familie möglicherweise nicht mehr hinter der Hilfe steht oder die Belastungen so groß geworden sind, dass eine auf den jungen Menschen fokussierte Hilfe nicht mehr erfolgversprechend ist.

4.9.7 Im Einzelnen: Betreuungshilfe / Erziehungsbeistand (Leistungstyp 5)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	16,5
Geschlecht m:w	63:37
durchschnittliche Hilfedauer (Monate)	7,8
Abbruchquote	22,9%
Anteil Folgehilfen HzE	20%
Anteil positiver Zielerreichung	59,3%
Anteil misslungener Hilfen	15,6%
Positive Entwicklungen in Familie	57%
Positive Entwicklungen beim jungen Menschen	72%
Negativentwicklungen in Familie	27%
Negativentwicklung beim jungen Menschen	16%
Effektstärke bei Familienzielen	gering bis mittel
Effektstärke bei Zielen junger Menschen	mittel bis hoch
Effektstärke bei Integrationszielen	hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	6
Anteil misslungener Hilfen	48,9%
Effektstärke bei Familienzielen	gering
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	9
Anteil positiver Zielerreichung	73,6%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

Betreuungshilfe / Erziehungsbeistandschaft zeigt eine hohe Wirksamkeit bei einer recht guten Stabilität der Hilfe. Die Hilfen richten sich an ein Klientel in einem relativ schwierigen Entwicklungsalter, dauert durchschnittlich nur 8 Monate und weist einen geringen Teil an Folgehilfen aus. Insofern: kein Handlungsbedarf.

4.9.8 Im Einzelnen: sozialpädagogische Familienhilfe (Leistungstyp 6)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	/
Geschlecht m:w	/
durchschnittliche Hilfedauer (Monate)	12
Abbruchquote	25,9%
Anteil Folgehilfen HzE	27,7
Anteil positiver Zielerreichung	56,9%
Anteil misslungener Hilfen	17,8%
Positive Entwicklungen in Familie	62%
Positive Entwicklungen beim jungen Menschen	64%
Negativentwicklungen in Familie	24%
Negativentwicklung beim jungen Menschen	21%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	6,9
Anteil misslungener Hilfen	52,2%
Effektstärke bei Familienzielen	keine
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	12
Anteil positiver Zielerreichung	70,2%
Effektstärke bei Familienzielen	hoch
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

Sozialpädagogische Familienhilfe hat eine hohe Wirksamkeit, etwas gemindert durch den Abbruch bei einem Viertel der Hilfen. Bei diesen fällt auf, dass insbesondere die schulische Entwicklung negativ verläuft. Eine Vermutung wäre, dass dann, wenn der schulische Problemdruck nicht nachlässt, die Familie keinen Sinn mehr in der Hilfe sieht. Zu Überlegen wäre, bei schulischen Zieldimensionen spezielle schulische Hilfen zu ergänzen, weil SPFH allein offensichtlich nicht gut geeignet ist, spezifische Schulprobleme zu lösen. Wenn dieses den Anteil der Abbrüche mindern könnte, würde die Durchschnittseffektivität ansteigen.

4.9.9 Im Einzelnen: ambulante ISE (Leistungstyp 7)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	15,1
Geschlecht m:w	58:42
durchschnittliche Hilfedauer (Monate)	9,6
Abbruchquote	26,9%
Anteil Folgehilfen HzE	38,4
Anteil positiver Zielerreichung	42,3%
Anteil misslungener Hilfen	23%
Positive Entwicklungen in Familie	42%
Positive Entwicklungen beim jungen Menschen	62%
Negativentwicklungen in Familie	37%
Negativentwicklung beim jungen Menschen	19%
Effektstärke bei Familienzielen	sehr negativ
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	gering
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	zu geringe Fallzahl
Anteil misslungener Hilfen	
Effektstärke bei Familienzielen	
Effektstärke bei Zielen junger Menschen	
Effektstärke bei Integrationszielen	
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	zu geringe Fallzahl
Anteil positiver Zielerreichung	
Effektstärke bei Familienzielen	
Effektstärke bei Zielen junger Menschen	
Effektstärke bei Integrationszielen	

Zusammenfassung und Hinweise:

Bei der geringen Fallzahl von 26 beendeten Hilfen können keine differenzierten Auswertungen gemacht werden.

4.9.10 Teilstationäre Hilfen in Tagesgruppen (Leistungstyp 8)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	11
Geschlecht m:w	79:21
durchschnittliche Hilfedauer (Monate)	15,4
Abbruchquote	40,4%
Anteil Folgehilfen HzE	31,7%
Anteil positiver Zielerreichung	43,4%
Anteil misslungener Hilfen	30,1%
Positive Entwicklungen in Familie	50%
Positive Entwicklungen beim jungen Menschen	56%
Negativentwicklungen in Familie	35%
Negativentwicklung beim jungen Menschen	34%
Effektstärke bei Familienzielen	keine
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	mittel
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	8,1
Anteil misslungener Hilfen	63,2%
Effektstärke bei Familienzielen	negativ
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	21
Anteil positiver Zielerreichung	69,8%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	sehr hoch
Effektstärke bei Integrationszielen	hoch

Wegen der hohen Abbruchquote sollen die Ergebnisse in den Tagesgruppen noch einmal genauer analysiert werden. Im Zeitraum 2011 und 2012 wurden von den fallverantwortlichen Fachkräften der Jugendämter die Wirkungen von 174 teilstationären Erziehungshilfen in Tagesgruppen (Berliner Leistungstyp 8) eingeschätzt. Das sind 36% aller beendeten erzieherischen Hilfen, die in Tagesgruppen erbracht wurden. Der Erfassungsanteil ist zwar klein, reicht für die Ermittlung repräsentativer Wirkungskennzahlen aber aus.

40,4%⁷ dieser Hilfen wurden nach Angaben der fallverantwortlichen SozialarbeiterInnen abgebrochen, also vorzeitig im Dissens beendet. Die Abbruchquote variiert zwischen den Bezirken ziemlich stark von 31 % bis 56,8 %. Insgesamt sind durchschnittlich 4 von 10 Hilfen nicht so durchgeführt worden, wie sie anfänglich geplant worden waren. In unserer Benchmark-Vergleichsgruppe von 707 beendeten Hilfen in Tagesgruppen liegt die Abbruchquote nur bei 24,2 %, also statistisch signifikant und mit einer bedeutenden Differenz niedriger als in Berlin. Die Berliner teilstationäre Abbruchquote liegt in der gleichen Größenordnung wie bei den Berliner stationären Hilfen.

43,4 % aller untersuchten Tagesgruppenunterbringungen weisen eine eindeutige Zielerreichung aus (Abbildung 18). In 30,1 % der abgeschlossenen Fälle wird dieses verneint. Die Effekte bei der Förderung der Kompetenz junger Menschen sind durchschnittlich gut, auch bei den Integrationsleistungen schneiden die Berliner Hilfen gut ab, die Effekte in der Familie sind aber durchschnittlich gleich Null (Abbildung 19). In der Benchmark-Vergleichsgruppe (Abbildung 20) sind die Durchschnittseffekte bei den Kompetenzen der jungen Menschen und deren Integration allerdings sehr hoch, die Effekte in den Familien sind gut bis sehr gut. In 53 % der Fälle gibt es eine eindeutige positive Zielerreichung. Insofern ist das relativ schlechte Abschneiden der Tagesgruppen in Berlin erkläungsbedürftig.

Teilstationäre Hilfen brauchen offensichtlich eine gewisse Mindestdauer. Bundesweit liegt die mittlere Hilfe-dauer bei 21,8 Monaten (Median), in Berlin dagegen bei 15,4 Monaten. Viele Hilfen brechen in der zweiten Jahreshälfte ab (durchschnittliche Dauer abgebrochener Hilfen: 8 - 9 Monate).

Abbildung 18: Zielerreichung von Hilfen in Tagesgruppen

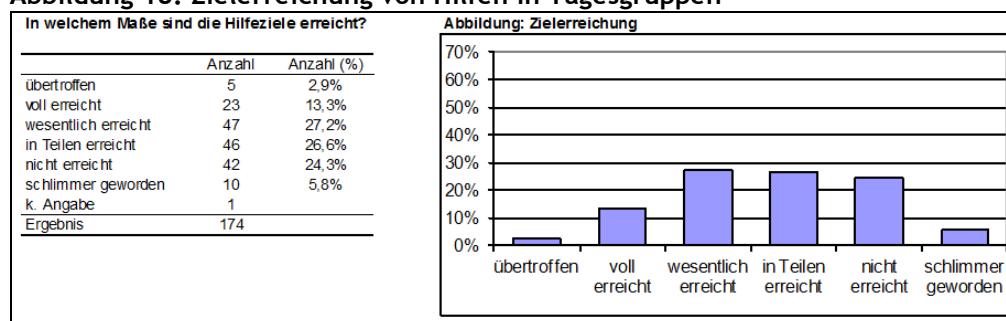
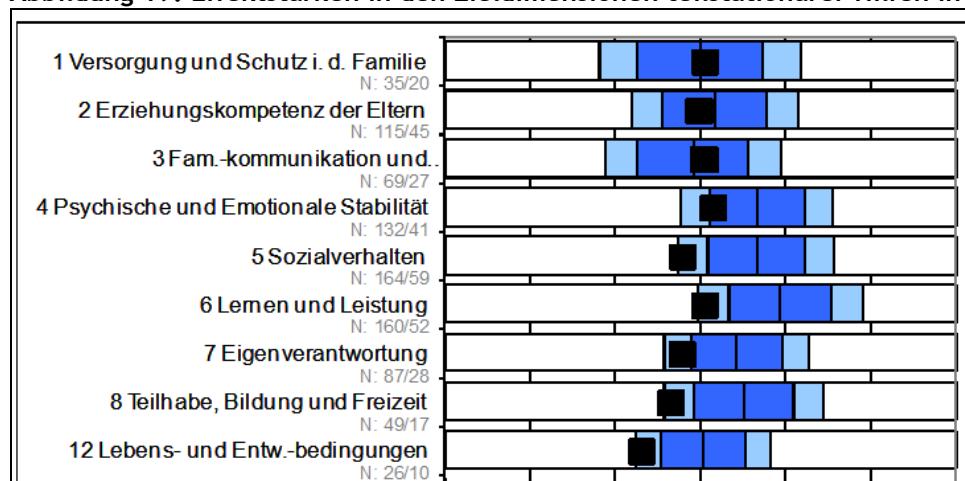
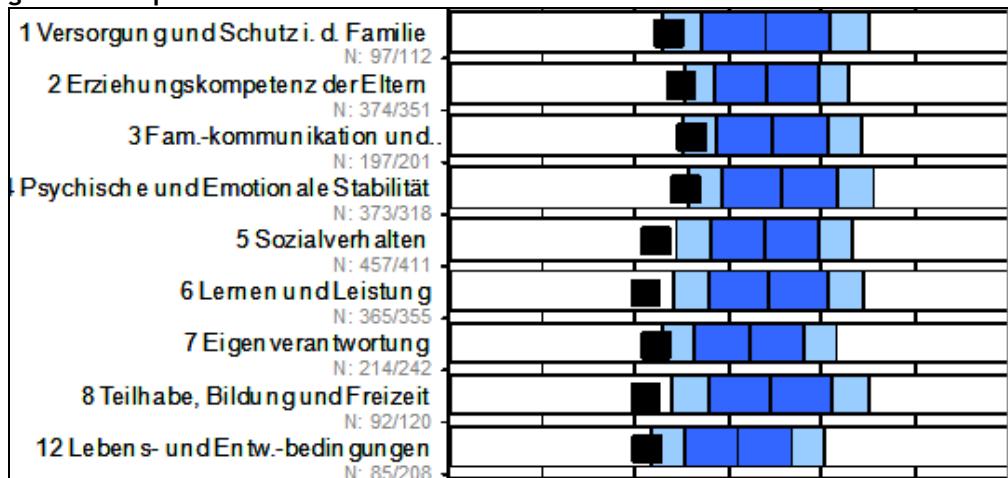


Abbildung 19: Effektstärken in den Zieldimensionen teilstationärer Hilfen in den fünf Berliner Bezirken



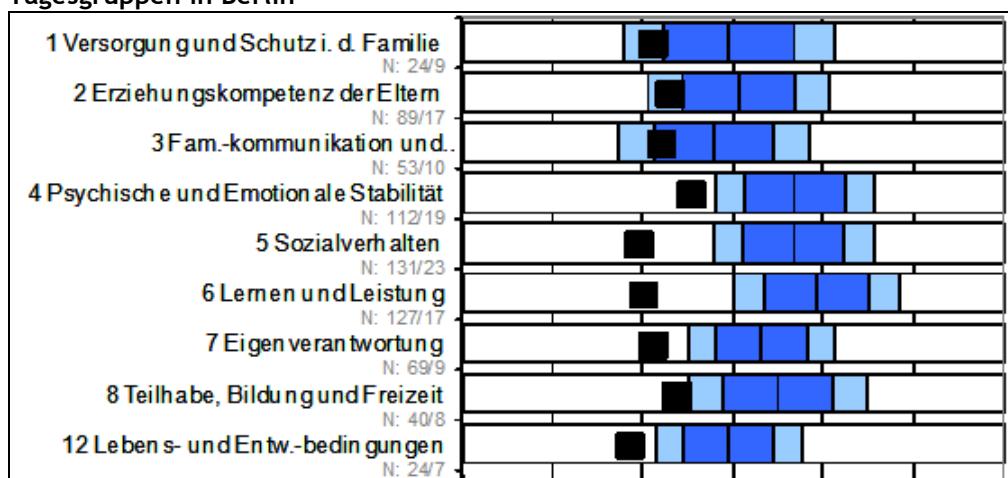
⁷ Der wahre Wert, den man findet, wenn alle Fälle dokumentiert worden wären, läge zwischen 33,1% und 47,7% (Konfidenzintervall p<0,05)

Abbildung 20: Effektstärken in den Zieldimensionen teilstationärer Hilfen in der nicht-Berliner Vergleichsstichprobe



Im Wesentlichen sind diese schlechteren Ergebnisse auf die hohe Abbruchquote zurückzuführen. Wenn Hilfen wie geplant bis zum Ende durchgeführt werden, ist auch in Berlin die Effektivität gut bis sehr gut (Abbildung 21).

Abbildung 21: Effektstärken bei den Zieldimensionen nicht-abgebrochenen teilstationären Hilfen in Tagesgruppen in Berlin



Ursachen für Abbrüche teilstationärer Hilfen

Auffällig ist das deutlich höhere durchschnittliche Aufnahmealter bei Tagesgruppen in Berlin (Median: 10,4 Jahre; Mittelwert: 10,8 Jahre / Benchmark-Vergleichsgruppe der Nichtberliner Tagesgruppen: Median und Mittelwert: 9,7 Jahre / dieser Unterschied ist statistisch hoch signifikant ($p < 0,001$)).

Es gibt einen signifikanten Zusammenhang zwischen dem Abbruchrisiko und dem Aufnahmealter. In Abbildung 22 wird eine Gruppierung der Klientel nach dem Aufnahmealter vorgenommen. Bei den Berliner Tagesgruppenfällen zeigt sich ein signifikant deutlich höheres Abbruchrisiko bei den über 12-Jährigen. Ähnlich stellt sich das bei den nicht-Berliner Fällen (Benchmark) dar (Abbildung 23), allerdings in viel geringerem Ausmaß.

Abbildung 22: Abbruchrisiko in Berliner Tagesgruppen in Abhängigkeit vom Aufnahmealter (N = 185)

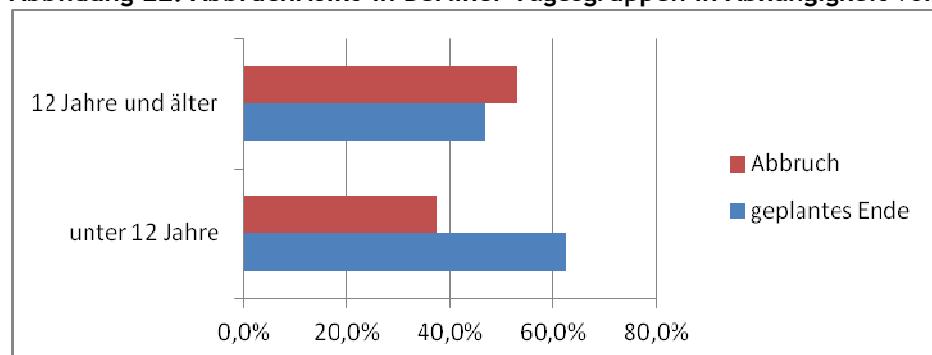
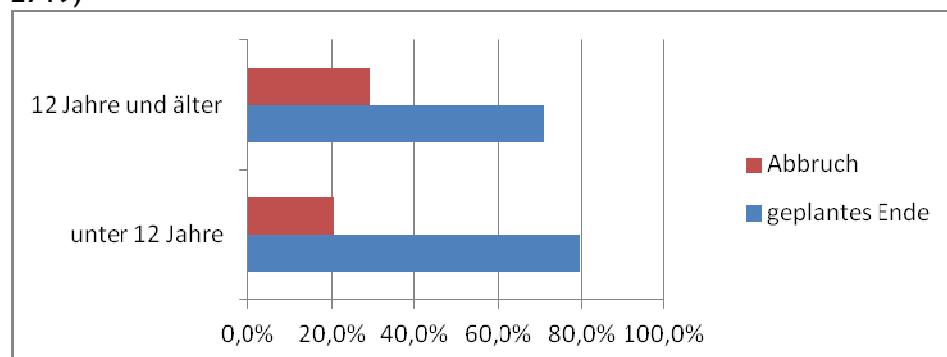


Abbildung 23: Abbruchrisiko in nicht-Berliner Tagesgruppen in Abhängigkeit vom Aufnahmealter (N = 2719)



Der Alterseinfluss reicht nicht aus, um das hohe Abbruchrisiko in Berlin zu erklären. 53,2 % der Hilfen bei den über 12-Jährigen (zum Zeitpunkt der Aufnahme) brechen in Berlin ab. In der Benchmarking-Stichprobe sind das nur 29 % der über 12-jährigen.

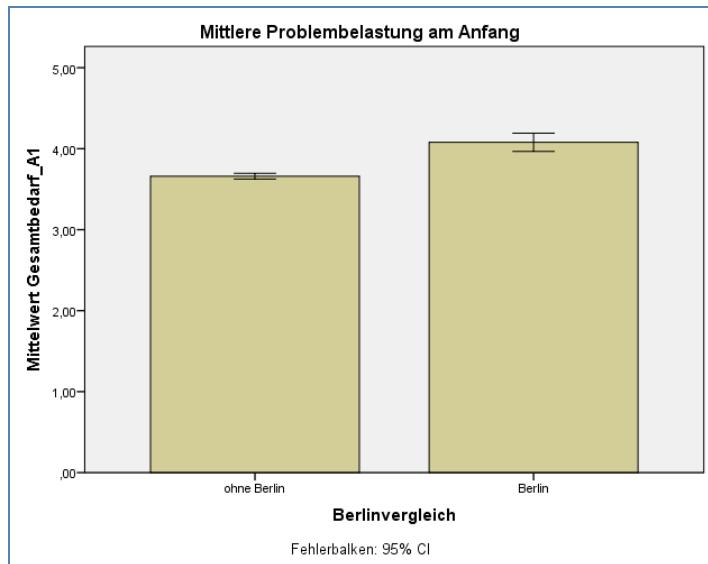
Tagesgruppen werden in Berlin auffällig seltener eingesetzt als im übrigen Bundesgebiet. Die Hilfen nach dem Leistungstyp 8 machen in Berlin einen Anteil von ungefähr 5 % aller Neufälle aus. In den nicht-Berliner Einrichtungen und Jugendämtern der Benchmarking-Stichprobe finden wir einen Anteil von 13,5 %. In der Bundesstatistik werden 20 % der teilstationären Hilfen in Tagesgruppen ausgewiesen (Destatis 2011: Anteil teilstationärer Hilfen an allen begonnenen Erziehungshilfen ohne Vollzeitpflege).

Dieser Befund lässt vermuten, dass die Indikation für die Tagesgruppenangebote in Berlin möglicherweise eine andere ist als im übrigen Bundesgebiet. Deswegen prüfen wir die Hypothese, dass sich die Tagesgruppenklientel in Berlin hinsichtlich der Bedarfslagen von anderen Regionen unterscheidet. So könnte es zum Beispiel sein, dass durch einen Selektionsmechanismus in Berlin die Kinder und Jugendlichen, die für eine Tagesgruppenarbeit anstehen, schwieriger sind oder andere Arten von Schwierigkeiten haben. Dieses wollen wir in der Folge überprüfen, wobei beachtet werden muss, dass die Daten aus der WIMES-Stichprobe nur Angaben über die Problembelastungen machen, nicht aber über die Gesamtkonstellation des Falles.

Vergleichen wir zunächst die Eingangs-Problembelastung:

- Die Gesamtbelastung der Berliner Klientel ist in der Tat signifikant um 0,4 Skalenpunkte höher (Abbildung 24).
- Die Unterschiede kommen hauptsächlich dadurch zustande, dass die Kinder auffälliger sind. Der Hauptunterschied liegt bei den schulischen Auffälligkeiten / Defiziten (Abbildung 25).
- Die Problembelastungen der Berliner Familien unterscheiden sich nicht von den nicht-Berliner Familien.

Abbildung 24: Unterschiede in den anfänglichen Problembelastungen bei Tagesgruppenklientel, differenziert nach Berlin und nicht-Berlin



Nicht jeder Bedarf muss automatisch Ziel der Veränderungsbemühungen sein. Dazu muss zuvor geklärt werden, welche Richtungsziele mit der Hilfe verfolgt werden sollen, welche Zielvereinbarungen mit den Klienten möglich sind und wohin deren Veränderungswünsche gehen. Auffällig ist bei den Berliner Hilfen, dass insgesamt weniger Richtungsziele vereinbart werden und dass sich weniger Ziele auf die Veränderung der familiären Situation beziehen. In den nicht-Berliner Hilfen ist es dagegen fast charakteristisch für die teilstationären Hilfen, dass sie sich gleichermaßen auf die Familie beziehen wie auf die Entwicklung der Kinder und Jugendlichen. Abbildung 26 zeigt, bei welchem Anteil der Fälle die WIMES-Dimension hilferelevant ist. Insgesamt werden in Berlin durchschnittlich weniger Richtungsziele vereinbart (Mittelwerte: 3,5 versus 4,4 Richtungsziele). Dieses kommt hauptsächlich dadurch zustande, dass in doppelt so vielen Fällen als bei den Nicht-Berliner gar keine Richtungsziele benannt werden (Abbildung 27).

Abbildung 25: Bedarfsprofile am Anfang der teilstationären Hilfen

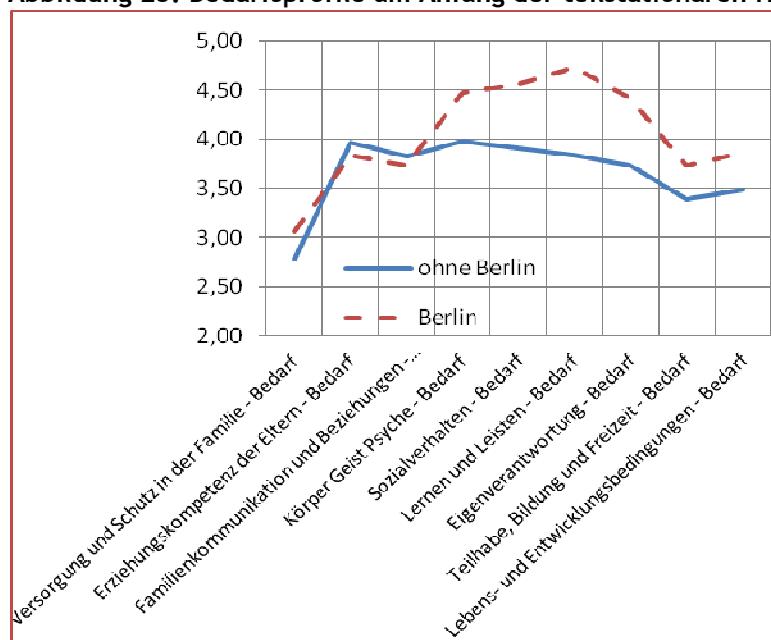


Abbildung 26: Anteil der als für die teilstationäre Hilfen relevant eingestuften Dimensionen; Vergleich zwischen Berlin und nicht-Berlin

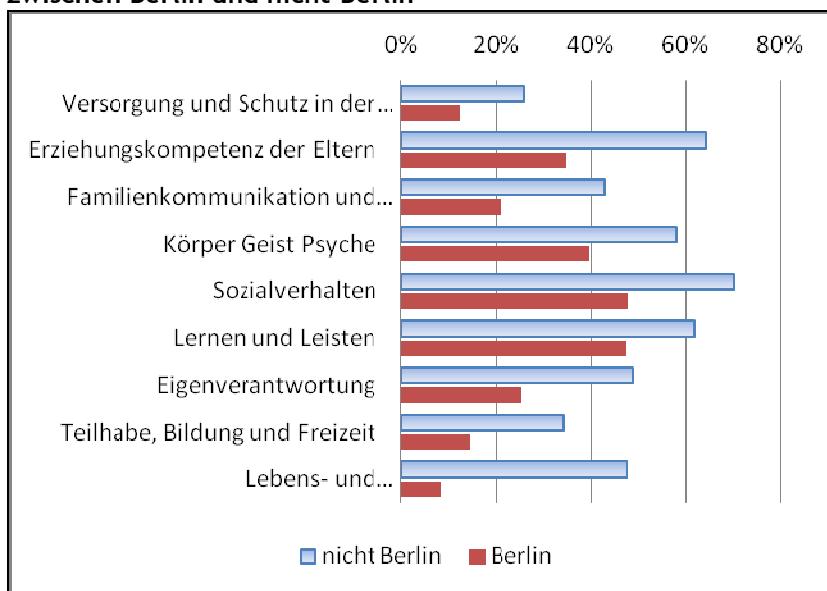
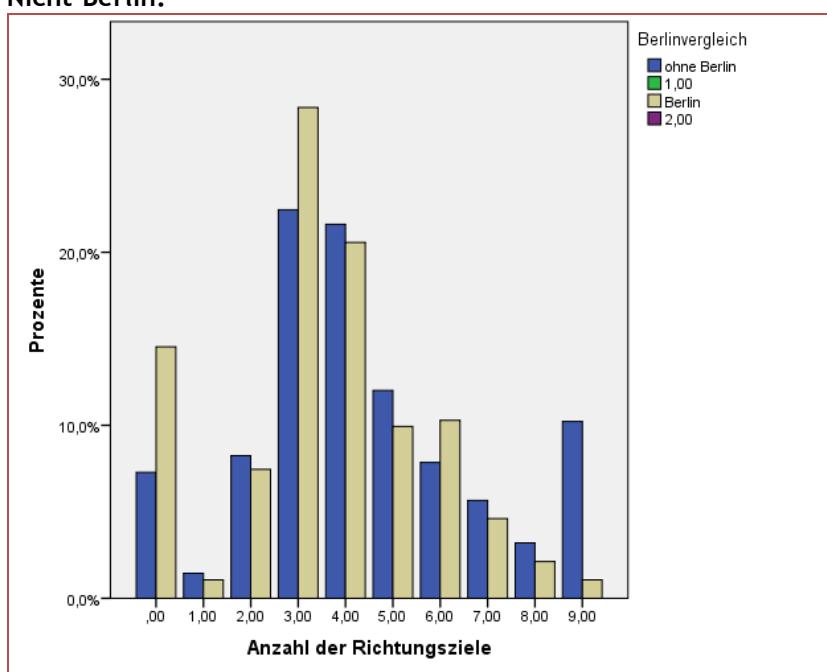


Abbildung 27: Anzahl (%) der Zieldimensionen teilstationärer Hilfen im Vergleich zwischen Berlin und Nicht-Berlin.



Zusammenfassung und Hinweise:

Zusammenfassend bleibt festzustellen, dass bei den Berliner teilstationären Hilfen im Vergleich zu Hilfen in nicht-Berliner Tagesgruppen...

- die Hilfeform seltener eingesetzt wird
- die jungen Menschen bei der Aufnahme etwas älter sind,
- die jungen Menschen auffälliger und problembelasteter sind, insbesondere im Bereich der Schule, im Sozialverhalten und in der Eigenverantwortung für die alltäglichen Dinge des Lebens,

- die Hilfen weniger Richtungsziele ausweisen und der Anteil der Hilfen, die gar keine Richtungsziele ausweisen, höher ist,
- die Zielsetzungen stärker auf die Entwicklungsaufgaben der jungen Menschen und weniger auf die Familie ausgerichtet sind.

Es sollte untersucht werden, warum Erziehungshilfe in Tagesgruppen in den fünf Bezirken auffällig seltener eingesetzt wird. Vielleicht gibt es keine entsprechenden Angebote. Vielleicht sind Tagesgruppen auch nicht effektiv genug, um eine Alternative zu stationären Unterbringungen darzustellen. Oder die Indikation für diese Hilfeart ist noch nicht genau definiert.

Erfolgreiche Tagesgruppen umfassen in der Regel ein schulisches und ein familienorientiertes Angebot. Durch die umfassende Inklusion von Förderschülern mit dem Förderschwerpunkt Soziales und Emotionales in die Regelschulen in Berlin könnten durch die Nicht-Tragbarkeit von Schülern im Regelsystem Bedarfe entstehen, die relativ teuer stationär im Berliner Umland untergebracht werden müssen, weil nur dort eine angemessene schulische Förderung gesichert ist. Dieser Bedarf deutet sich bei der beschriebenen Tagesgruppen-Klientel in der Dimension 6 (Abbildung 25) an, auch in anderen Ländern macht man diese Erfahrung. Bevor also durch Schulprobleme die Fallzahl stationärer Hilfen erneut steigt, sollte man möglicherweise schulintegrierte teilstationäre Erziehungshilfen ausbauen. Diese müssen aber auf alle Fälle schulfördernde und schulintegrierende Konzeptanteile haben und eine intensive Elternarbeit umfassen.

4.9.11 Stationäre Hilfen

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	14,4
Geschlecht m:w	54:46
durchschnittliche Hilfedauer (Monate)	12,5
Abbruchquote	42,9%
Anteil Folgehilfen HzE	50,8%
Anteil positiver Zielerreichung	47,3%
Anteil misslungener Hilfen	23,7%
Positive Entwicklungen in Familie	53%
Positive Entwicklungen beim jungen Menschen	66%
Negativentwicklungen in Familie	37%
Negativentwicklung beim jungen Menschen	22%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	mittel
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	9,3
Anteil misslungener Hilfen	52,6%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	14,3
Anteil positiver Zielerreichung	74,3%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	mittel bis hoch
Effektstärke bei Integrationszielen	hoch

Die Ergebnisse des Zwei-Jahres-Zeitraums bestätigen die Befunde und Empfehlungen aus dem ersten Projektbericht vom 10.02.2012, wo sie ausführlich diskutiert wurden. Die stationären Hilfen werden hier trotzdem noch einmal detailliert betrachtet, weil

- sie die intensivsten Eingriffe in die Biografie eines jungen Menschen und seiner Familie sind,
- die Zielgruppen am gefährdetsten sind, langfristig und möglicherweise über Generationen auffällig zu sein und
- sie die teuersten Hilfen sind.

Die Strategie, diese Hilfen zu vermeiden und allein auf präventive und sozialräumliche Alternativen zu setzen, greift zu kurz. Die Anzahl der stationären Hilfen ist seit Jahrzehnten relativ konstant und es wird diese

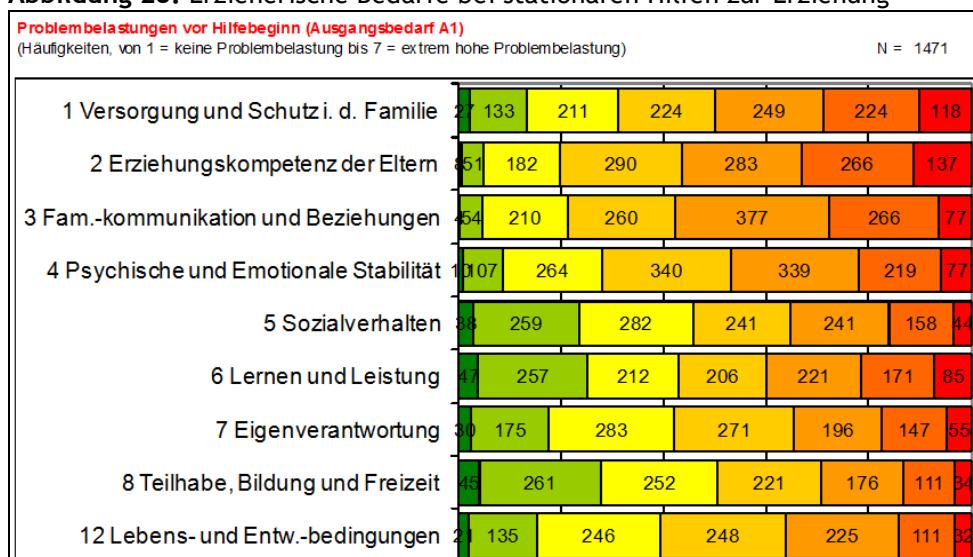
Hilfeform auf absehbare Zeit geben. Umso wichtiger sind die Bemühungen, stationäre Angebote möglichst effektiv und effizient zu machen. Niedrige Entgelte und kurze Hilfedauer als Kostenstrategien greifen ebenfalls zu kurz, weil die Angebote möglicherweise qualitativ nicht ausreichen und nicht gelingende Hilfen zu Langzeitfolgen mit hohen Gemeinschaftskosten führen (wenngleich oft außerhalb der Jugendhilfe). Es geht kein Weg vorbei an einer Strategie der Effektivierung stationärer Hilfen bei leistungsgerechten Kosten.

In den 5 Berliner Bezirken werden ungefähr 28 % der begonnenen Hilfen stationär durchgeführt, wenn man bei der Bezugsgröße aller Hilfen die therapeutischen Hilfen mitzählt. Das liegt weit unter dem Bundesdurchschnitt, ein Vergleich, der allerdings einer Verzerrung unterliegt, weil in anderen Jugendämtern nicht im Berliner Umfang therapeutische Hilfen geleistet werden. Ohne Berücksichtigung der therapeutischen Hilfen liegt der Anteil bei 33 %. Das entspricht Anteil in der Benchmark-Stichprobe. In 27 % der Hilfen werden in intensiven Settings durchgeführt. Dieser Anteil variiert zwischen den Bezirken von 21 % bis 44 %.

Das Aufnahmealter mit einem Median von 14,4 Jahren entspricht ungefähr dem Benchmark. Es werden etwas mehr Jungen als Mädchen stationär untergebracht, wobei das Aufnahmealter bei Mädchen im Durchschnitt (Median) ein Dreivierteljahr über dem der männlichen Klienten liegt. Diese auch in der Benchmark-Stichprobe anzutreffende Disparität wird unterschiedlich bewertet und erklärt. Zum einen zeigen Mädchen ihre Probleme auf andere Weise, weniger „acting-out“ und mehr auf eine stille, möglicherweise selbstschädigende Art - dieses scheint sich aber in den letzten Jahren zu ändern, weil Gewalt und heftige Brutalität zunehmend bei Mädchen vorkommen. Ein anderer Grund liegt wohl darin, dass Mädchen besser mit den schulischen Anforderungen klarkommen, aber in der Pubertät sehr viel emotionaler in der Familie an ihrer Ablösung arbeiten. So können Mädchen mitunter jahrelang unauffällig sein, um plötzlich in kurzer Zeit in heftige Konflikte mit ihren Eltern und generell mit Erwachsenen zu geraten. Auch ist die Morbidität psychischer Erkrankungen bei Mädchen in der Pubertät sehr hoch.

Die stärksten Bedarfe zeigen sich bei der stationären Klientel in den Dimensionen Erziehungskompetenz / Familiensystem (ungefähr 70 % erhebliche bis extreme Problembelastungen) und bei den psychischen Auffälligkeiten der Kinder und Jugendlichen (60 %). In 60 % der Fälle sind die Versorgung und der Schutz in der Familie in erheblicher Weise nicht gewährleistet. Vergleichsweise weniger extrem problematisch sind das Sozialverhalten und die Integration (45 %).

Abbildung 28: Erzieherische Bedarfe bei stationären Hilfen zur Erziehung



Die meisten stationären Hilfen zielen auf die Verselbständigung junger Menschen ab (43%), haben also nicht mehr das Ziel einer Reintegration in die Herkunftsfamilie. Letzteres wird nur bei 27% der jungen Menschen angestrebt. Lediglich in einem Bezirk ist der Anteil der jungen Menschen größer, die nach einer stationären Hilfe wieder in die Familien zurückkehren sollen. Entsprechend jünger sind dort die Kinder bei der Aufnahme.

In 34% der begonnenen Hilfen hat unmittelbar davor bereits eine andere stationäre Hilfe stattgefunden⁸. Dieser Anteil entspricht dem der Benchmark-Stichprobe und weist auf ein Problem der Heimerziehung hin: Viele Kinder und Jugendliche sind - manchmal unterbrochen durch Zwischenaufenthalte zu Hause, auf der Straße, in der Psychiatrie usw. - auf dem Weg durch eine Reihe von Heimen und sonstigen Wohnformen.

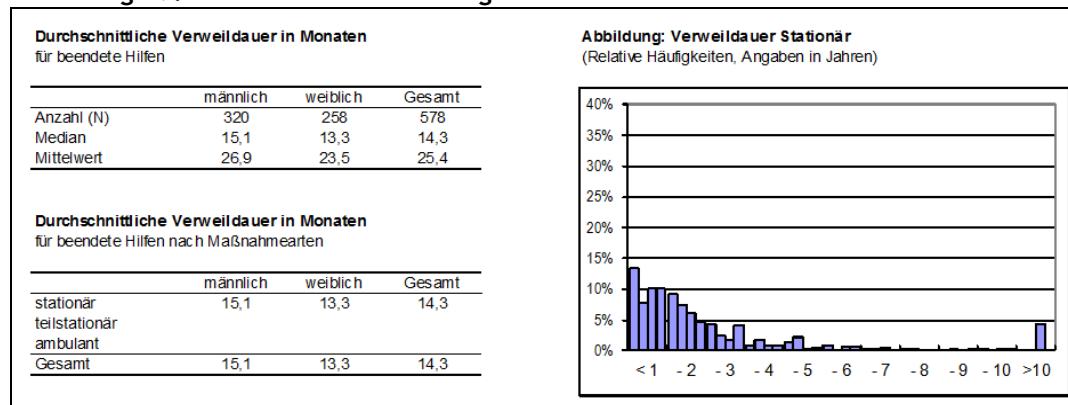
Die Durchschnittsdauer der stationären Hilfen ist ein nicht eindeutiger Qualitätsindikator. Eine kurze Dauer kann bedeuten, dass Ziele relativ schnell erreicht wurden, dass der Hilfesort oder die Hilfeform bei Bedarf frühzeitig angepasst wurden oder dass nicht passende Hilfen rechtzeitig beendet wurden. Genauso wenig kann eine lange Hilfedauer als Qualitätsmerkmal genutzt werden, beispielsweise dann nicht, wenn eine Hilfe zur Reintegration in die Familie gedacht war. In jedem Fall muss man die Zielsetzung und das Ergebnis zur Dauer in Beziehung setzen.

In den 5 Bezirken dauert eine stationäre nicht abgebrochene Hilfe im Median⁹ 14,3 Monate (Abbildung 29). Das sind 3,5 Monate weniger als in der Benchmarking-Stichprobe.

Stationäre Hilfen sind trotz der großen Problembelastungen der jungen Menschen und problematischer Lebensverhältnisse sehr erfolgreich. Wir müssen allerdings einschränken: Sie sind nur dann wirklich erfolgreich, wenn sie wie geplant bis zum vereinbarten Ende durchgeführt werden. Abbildung 30 zeigt die Effektstärken in den neun WIMES-Dimensionen (Richtungsziele). Am wirksamsten wird die psychische Stabilität und Gesundheit gefördert und eine soziale Integration erreicht. Auch die familiäre Situation kann deutlich verbessert werden. Entsprechendes zeigt sich in der rückblickenden Entwicklungseinschätzung durch die Fallverantwortlichen (Abbildung 31): Dreiviertel der jungen Menschen machen eine positive bis sehr positive Entwicklung. Die Entwicklung in den Familien wird allerdings nicht so positiv eingeschätzt.

Insgesamt wird in 70 % der nicht abgebrochenen Hilfen das Ziel ohne Abstriche als erreicht angesehen (Abbildung 32). Nur in 6,5 % der Fälle sehen die Fachleute die Hilfe als gescheitert an. Das entspricht bis aufs Komma genau dem Ergebnis in der Benchmarking-Stichprobe.

Abbildung 29: Hilfedauer bei nicht abgebrochenen stationären Hilfen



⁸ . Wir haben nicht erhoben, ob es in der Biographie junger Menschen früher schon einmal eine Heimunterbringung gegeben hatte, sondern nur nach den unmittelbaren Hilfesfolgen gefragt

⁹ Der arithmetische Mittelwert sollte nicht als Indikator verwendet werden, weil die Verteilung linksschief ist und bereits wenige Langzeitfälle den Mittelwert nach oben ziehen.

Abbildung 30: Effektstärken bei stationären nicht abgebrochenen Hilfen (nur Richtungsziele)

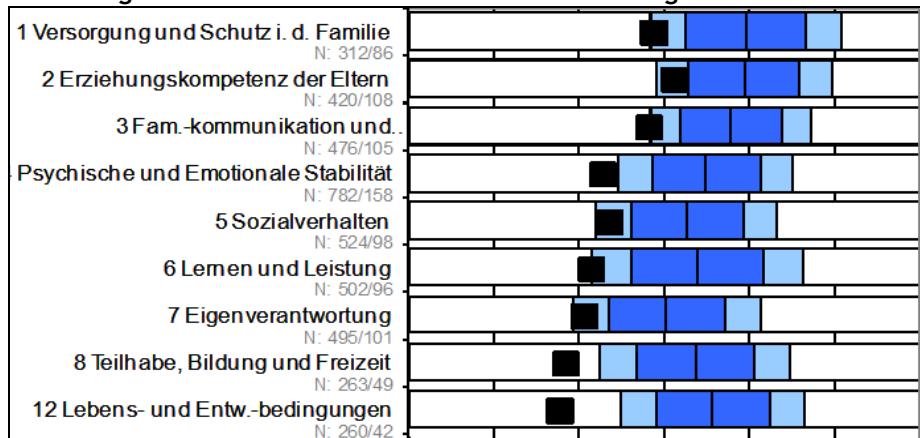


Abbildung 31: Retrospektive Einschätzung der Entwicklung während nicht abgebrochener stationärer Hilfen

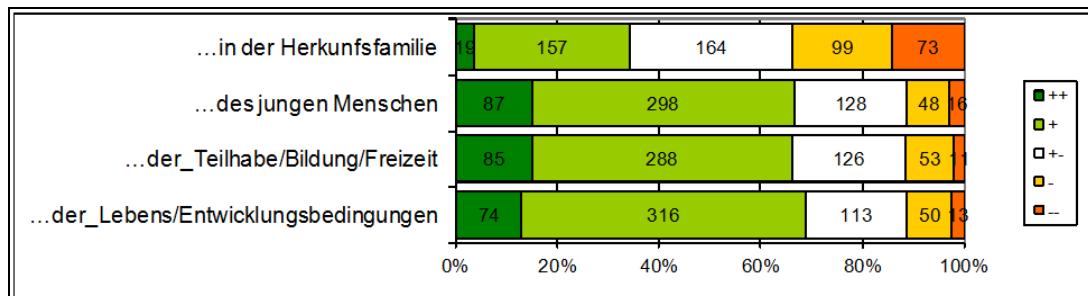
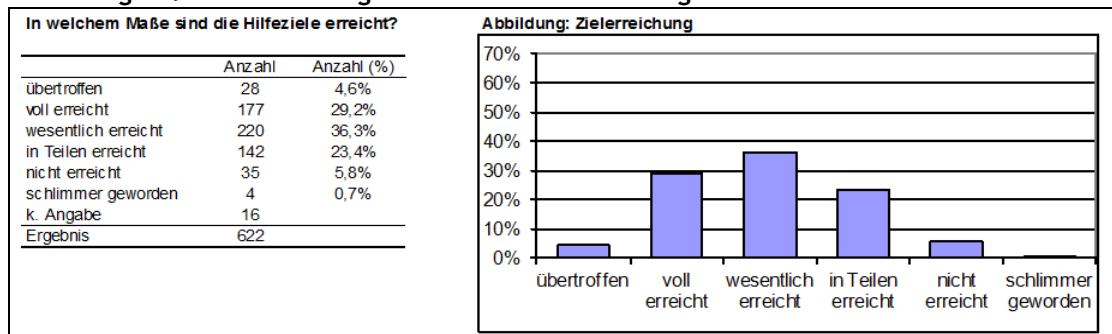
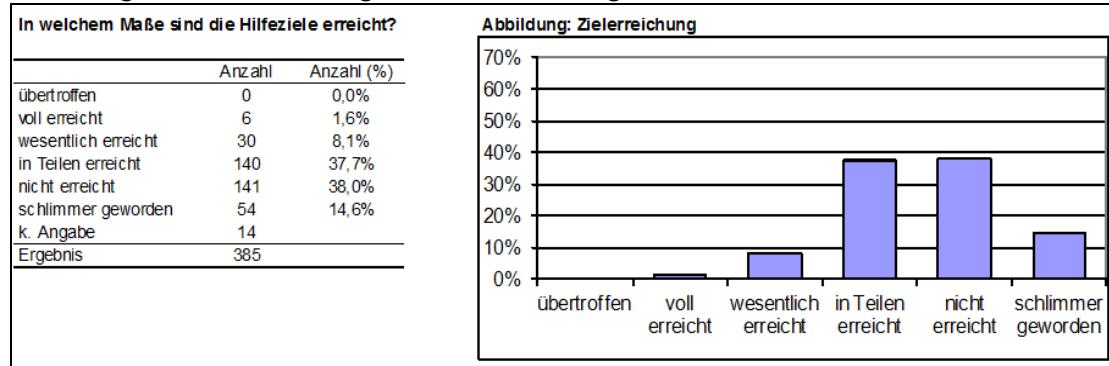


Abbildung 32: Zielerreichung bei stationären nicht abgebrochenen Hilfen



Alle könnten also mit dieser Hilfeform zufrieden sein, wenn es nicht die hohe Zahl der vorzeitig durch Abbruch beendeten Hilfen gäbe. Denn genauso eindeutig, wie die stabil durchgeföhrten Hilfen erfolgreich sind, so eindeutig sind die abgebrochenen Hilfen in der Regel ein Misserfolg. Stellen wir gleich einmal die Zielerreichung gegenüber (Abbildung 33). Nur 10 % der Hilfen haben ihre Ziele erreicht. In 53 % der Fälle war die Intervention ein Misserfolg.

Abbildung 33: Zielerreichung bei stationären abgebrochenen Hilfen

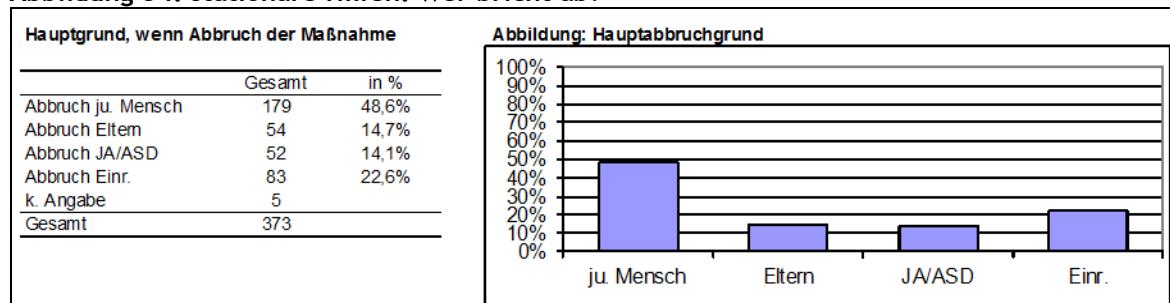


Ein Problem bei den stationären Hilfen entsteht also daraus, dass der Anteil der Abbrüche so hoch ist. Bei der Mischung der beiden Teilgruppen aus Abbruch und Nicht-Abbruch zeigt sich das Gesamtbild einer nur mäßigen Wirksamkeit.

Umso dringender stellt sich die Frage, durch welches fachliche Vorgehen und unter welchen Rahmenbedingungen die Abbruchrisiken gesenkt werden können. Dieses lässt sich durch Evaluation der Wirkungen nicht lösen. Dazu bedarf es einer wissenschaftlichen Ursachenforschung und fachlicher Dialoge. Solche Untersuchungen liegen vor, aber noch nicht in ausreichendem Maße. Außerdem sollten örtliche Besonderheiten berücksichtigt werden, weil - wie gesagt - die Abbruchquoten von Jugendamt zu Jugendamt und von Einrichtung zu Einrichtung stark variieren. In der bundesweiten Untersuchung ABIE zu den Ursachen von stationären Abbrüchen, an der auch Berliner Fälle beteiligt waren (Tornow & Ziegler, 2012), konnten die Beendigungsarten zu 36% durch Merkmale der Organisations- und Leitungskultur in den Heimen und aus der Qualitätsorientierung in den Beziehungen zwischen Jugendamt und Leitungserbringer erklärt werden. Auch dieser Befund deutet darauf hin, dass ersten Abbruchquoten durch Organisations- und Konzeptentwicklung veränderbar sind und das zweitens die qualitätsorientierte Kooperation zwischen öffentlicher und freier Jugendhilfe erheblich Verbesserungspotenziale birgt.

Auch wenn ein Abbruch immer eine Koproduktion ist, sind es im Wesentlichen die Klienten, die zu Zweidrittel die Initiative ergreifen (Abbildung 34). Bemerkenswert ist allerdings, dass es immerhin durchschnittlich 9 Monate dauert, bis Hilfen abbrechen. Wenn man nicht miteinander klar kommt, wenn die anfängliche Hilfeplanentscheidung nicht die richtige war, wenn es kein Arbeitsbündnis gibt, sollte man das früher feststellen können. Nur 20 % der Abbrüche ereignen sich im ersten Quartal. Man sollte überlegen, ob die Zeit der Eingewöhnung und der Festigung und Präzisierung eines Arbeitsbündnisses nicht besser genutzt werden kann, die Indikation zu überprüfen und frühzeitig Planungen zu korrigieren oder die Hilfe proaktiv abzubrechen.

Abbildung 34: stationäre Hilfen: Wer bricht ab?



Betrachten wir die Abbruchquote noch etwas genauer. Insgesamt unterscheiden sich die 5 Projektbezirke in ihren stationären Abbruchquoten kaum. Anders sieht es allerdings aus, wenn man das Aufnahmealter berücksichtigt. Die Abbruchwahrscheinlichkeit hängt stark vom Eintrittsalter zu Beginn der Hilfe ab (Tabelle 7). Der Zusammenhang zwischen Aufnahmealter und Abbruchrisiko ist nicht linear: Ab 16,5 Jahren fällt die Abbruchwahrscheinlichkeit wieder ab. Die Abhängigkeit der Abbruchquote vom Eintrittsalter ist übrigens noch ausgeprägter, wenn man das Geschlecht des Klienten berücksichtigt: In der kritischen Altersgruppe von 12 bis 17 Jahren liegt das Abbruchrisiko bei Jungen bei 62,3 %(!).

Tabelle 7: Abbruchquote und Aufnahmealter bei stationären Hilfen¹⁰

Abbruchquote bei stationären Hilfen	Unter 12 Jahren	12 bis 16,5 Jahre	Über 16,5 Jahre	Gesamt
5 Bezirke gesamt	25,8%	58,9%	40,5%	43,1%
5 Bezirke im Einzelnen	23,6% - 26,7%	48,3% - 63%	36,8% - 51,4%	39,7% - 47,4%

Interessant dürfte noch die Dauer der stationären Hilfen sein. Da diese stark mit der Abbruchquote zusammenhängt, differenzieren wir die Auswertung nach diesen beiden Gruppen. Hilfen, die abbrechen, dauern durchschnittlich (Median) 9,4 Monate. Der Umstand, dass Hilfen früher abbrechen, ist eher positiv zu bewerten. Problematisch ist es, wenn sich Abbrüche erst nach Jahresfrist oder noch später ergeben.

Tabelle 8: Hilfedauer in Jahren (Median)

	5 Bezirke im einzelnen	Berlin 5 Bezirke insgesamt
Bei geplanten Beendigungen	1,08 bis 1,68	1,39
Bei Abbrüchen	0,47 bis 1,18	0,78
Gesamt	0,77 bis 1,36	1,10

Unabhängig von der Durchschnittsdauer wäre es wünschenswert, dass die Abbrüche, wenn es schon zu ihnen kommt, so früh wie möglich stattfinden. Insofern ist es ein Indikator für die Qualität der Hilfesteuerung, wenn die Differenz zwischen der Hilfedauer bei geplanten Beendigungen und bei Abbrüchen hoch ist. Dieses ist in der Betrachtung der Gesamtstichprobe der Fall. Dennoch ist die durchschnittliche Dauer der abgebrochenen Hilfen nach über 9 Monaten hoch, bedeutet es doch, dass sich 50 % der Abbrüche nach einem Dreivierteljahr und mehr ereignen.

Bei der Diskussion zum ersten Projektbericht (10.2.2012) wurde der (vielleicht nicht ganz ernst gemeinte) Vorschlag geäußert, dass man doch die Hälfte der (teuren) stationären Maßnahmen weglassen könnte, wenn sie ohnehin in einen Abbruch einmünden. Das wäre ein gar nicht so schlechter Vorschlag, wenn man denn vorher wüsste, mit welchem Frühindikator man das Scheitern vorhersehen könnte und wenn es denn eine Alternative gäbe. Der Bürger verliert ja nicht dadurch sein Recht auf eine Sozialleistung, dass diese mit hoher Wahrscheinlichkeit nicht wirkt. Der Anspruch ist nur vom Bedarf abhängig und nicht von Ursache und erwarteter Wirkung. Andererseits hat der Bürger ein Anrecht auf eine geeignete Leistung. Offensichtlich ist die Hälfte der Leistungen im Nachherein betrachtet nicht geeignet gewesen, nachhaltig zu Verbesserungen zu führen. Welcher Bürger würde sich wissentlich eine Dienstleistung einkaufen, die nur in jedem zweiten Fall ein Problem löst?

Zur Rechtfertigung der hohen Abbruchquote könnten Jugendämter und Leistungserbringer einwenden, dass Dienstleistungen immer Koproduktionen mit dem Nutzer seien, und dass ohne dessen Mittun, Geduld, Einsicht, Motivation usw. ein Erfolg nicht zu erreichen sei. Mit anderen Worten: Wenn es gut geklappt hat, liegt das an der Qualität der Fachlichkeit, wenn nicht, liegt das an dem mangelnden Willen oder gar dem Widerstand der Nutzer. Hier stellt sich die Frage, welche Verantwortung für das Produkt „Jugendhilfe“ denn der Produzent bereit ist zu übernehmen. Sicherlich ist es eine gute Entwicklung, dass der Bürger vor übergriffiger Fürsorge oder professioneller Besserwisserei geschützt ist und dass er nicht der Autonomie, sein Leben selber zu gestalten, beraubt wird. Wenn dieses aber dazu führt, dass der Dienstleister nicht mehr die Verantwortung für die Qualität seiner Angebote einschließlich der Ergebnisqualität übernehmen will, hat sich die emanzipatorische Entscheidung der Sozialen Arbeit in eine billige professionelle Selbstentlastungsstrategie verwandelt. Zur Professionalität gehört immer auch eine Ethik der Verantwortungsübernahme.

Zusammenfassung und Hinweise:

¹⁰ Die Gliederung in genau diese drei Altersgruppen wird durch die Ergebnisse des ersten Projektberichtes vom 10.2.2012 nahe gelegt. Mittels multivariater Analysen wird genau die Altersgruppe zwischen 12 und 16,6 Jahren als die risikoreichste ausgewiesen.

Stationäre Hilfen sind erfolgreich, wenn sie wie geplant zu einem ordentlichen Ende durchgeführt werden. Die hohe Abbruchquote senkt aber die Effektivität auf ein nur mäßiges Niveau. Abgebrochene Hilfen führen relativ oft zu weiteren Hilfen oder gar zu Hilfeketten, in denen jede zukünftige Hilfe eine geringere Erfolgswahrscheinlichkeit hat.

Die Effektivität und auch die Effizienz (nachhaltiger Kostenaspekt) kann erhöht werden, wenn

- Hilfen frühzeitig eingeleitet werden,
- geeignete Konzepte in der Altersgruppe 12-16 eingesetzt werden,
- Abbruchrisiken früh erkannt werden und in der Hilfeplanung darauf proaktiv reagiert wird,
- Fachleute lernen, mit Risikoabschätzungen und statistischem Wissen zu arbeiten (risk-literacy),
- andere, noch nicht erkannte Maßnahmen ergriffen werden.

Um die Ursachen und Rahmenbedingungen für hohe Abbruchrisiken zu kennen und daraus Strategien abzuleiten, bedarf es einer gezielten Qualitätsentwicklung in Berlin. Geeignete Maßnahmen können sein:

- Empirische Erforschung von Gelingens- versus Risikofaktoren durch eine Tiefenanalyse ausgewählter Fälle.
- Qualitätsdialoge in Form „Runder Tische“, Expertengruppen, Delphi Foren, E-Learning-Foren u. ä..
- Strategie-Workshops mit systemdynamischen Modellen (z.B. ABIE-Spiel).
- Praxis-Forschungsprojekte (wissenschaftlich begleitete Projekte alternativer Jugendhilfestrategien).
- Bewerbung um Teil an Projekten des Max-Planck-Instituts für Bildungsforschung in Berlin

Erkenntnisse können sich auswirken auf eine gezieltere psychosoziale Diagnostik, Auswahl geeigneter Angebote, andere Formen der Steuerung bei Risikogruppen, Schaffung geeigneter Angebote, intensivere Hilfebegleitung risikoreicher Hilfen. Hilfreich wären Instrumente zur Identifizierung abbruchriskanter Fälle und zum „monitoring“ der Abbruchswahrscheinlichkeiten im Hilfelebenslauf. Möglicherweise muss auch die Familie anders begleitet werden, weil zu erwarten ist, dass in vielen Hilfen die Haltung der Eltern zu der Hilfe sehr ambivalent.

4.9.12 Im Einzelnen: stationär (Gruppenangebot) (Leistungstyp 9)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	13,7
Geschlecht m:w	58:42
durchschnittliche Hilfedauer (Monate)	12,2
Abbruchquote	49,9% ¹¹
Anteil Folgehilfen HzE	49,8%
Anteil positiver Zielerreichung	39,5%
Anteil misslungener Hilfen	28,3%
Positive Entwicklungen in Familie	48%
Positive Entwicklungen beim jungen Menschen	62%
Negativentwicklungen in Familie	34%
Negativentwicklung beim jungen Menschen	24%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	mittel bis hoch
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	12
Anteil misslungener Hilfen	50,8%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	keine ¹²
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	12,9
Anteil positiver Zielerreichung	63,4%
Effektstärke bei Familienzielen	mittel
Effektstärke bei Zielen junger Menschen	mittel bis hoch
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

¹¹ bei den über 12-Jährigen liegt die Abbruchquote bei 60,7% (!)

¹² bis auf mittlere Effekte bei der psychischen Entwicklung

Hier gelten die gleichen Aussagen wie für die stationären Hilfen allgemein. Die Abbruchquote ist sogar noch einmal höher.

Bei diesem Leistungstypus sollte einmal geprüft werden, warum gerade hier die Erfassungsquote in der Evaluationsstichprobe so gering ist (40% der Beendigungen versus 53% im Durchschnitt). Eine Hypothese könnte sein, dass diese Hilfen nicht die Aufmerksamkeit genießen, die sie bräuchten. Möglicherweise gibt es aber gerade für diese Fälle ohnehin keine Alternative, so dass die Fallverantwortlichen froh um jeden Monat sind, den die Hilfe „hält“. Dafür würde auch sprechen, dass es keinen Unterschied in der Hilfedauer gibt, egal ob sie abbricht, oder geplant beendet wird.

Da bei diesen Hilfen in 40% der Fälle eine Rückkehr in die Familie geplant ist, sollte der Frage nachgegangen werden, wie die Entwicklung in der Familie begleitet wurde.

4.9.13 Im Einzelnen: stationär (Gruppenangebot Wohngemeinschaft) (Leistungstyp 10)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	16,4
Geschlecht m:w	53:47
durchschnittliche Hilfedauer (Monate)	12,5
Abbruchquote	55%
Anteil Folgehilfen HzE	39,3%
Anteil positiver Zielerreichung	39%
Anteil misslungener Hilfen	25%
Positive Entwicklungen in Familie	45%
Positive Entwicklungen beim jungen Menschen	60%
Negativentwicklungen in Familie	36%
Negativentwicklung beim jungen Menschen	27%
Effektstärke bei Familienzielen	keine bis mittel
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	keine
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	12,5
Anteil misslungener Hilfen	44,2
Effektstärke bei Familienzielen	gering bis mittel
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	18,6
Anteil positiver Zielerreichung	74,7
Effektstärke bei Familienzielen	mittel bis hoch
Effektstärke bei Zielen junger Menschen	hoch
Effektstärke bei Integrationszielen	sehr hoch

Zusammenfassung und Hinweise:

Hier gelten die gleichen Aussagen wie für die stationären Hilfen allgemein. Die Abbruchquote ist sogar noch einmal höher.

4.9.14 Im Einzelnen: stationär (familienanalog) (Leistungstyp 11)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	6,6
Geschlecht m:w	52:48
durchschnittliche Hilfedauer (Monate)	21
Abbruchquote	25,3%
Anteil Folgehilfen HzE	74,4%
Anteil positiver Zielerreichung	65%
Anteil misslungener Hilfen	10,8%
Positive Entwicklungen in Familie	44%
Positive Entwicklungen beim jungen Menschen	74%
Negativentwicklungen in Familie	40%
Negativentwicklung beim jungen Menschen	17%
Effektstärke bei Familienzielen	keine bis mittel
Effektstärke bei Zielen junger Menschen	gering
Effektstärke bei Integrationszielen	mittel
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	18,1
Anteil misslungener Hilfen	40%
Effektstärke bei Familienzielen	/
Effektstärke bei Zielen junger Menschen	negativ
Effektstärke bei Integrationszielen	keine
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	21
Anteil positiver Zielerreichung	75%
Effektstärke bei Familienzielen	keine bis mittel
Effektstärke bei Zielen junger Menschen	gering bis mittel
Effektstärke bei Integrationszielen	gering

Zusammenfassung und Hinweise:

Familienanaloge Hilfen nehmen unter den stationären Hilfen eine Sonderstellung ein. Sie dauern länger, brechen nicht auffällig oft ab und haben in der Regel positive Ergebnisse. Diese spiegeln sich aber nicht den WIMES-Dimensionen, hauptsächlich wohl deswegen, weil nicht die Verhaltensprobleme im Vordergrund stehen, sondern die geeignete dauerhafte Unterbringung. Bei längeren Hilfen muss die Wirkungsevaluation im Verlauf fortgeschrieben werden, weil die Bedarfslagen sich mit der Entwicklung des Kindes ändern.

Wenn Verhaltensauffälligkeiten im Hilfesetting zunehmen, kommt es eher zu einem Abbruch der Hilfe. Welches dafür wiederum die Ursachen sind, müsste erforscht werden. Hypothesen könnten sein: Das Kind fühlt sich nicht wohl. Die Kinder sind für ein familienanaloges Setting zu schwierig (Störungen des Sozialverhaltens,

Bindungsstörungen, andere psychische Beeinträchtigungen). Auch Ambivalenzen in der Herkunftsfamilie können den Entwicklungsprozess beeinträchtigen.

4.9.15 Im Einzelnen: stationär (Individualangebot / BEW) (Leistungstyp 12)

Kennzahl (alle Hilfen)	Befund
Alter bei Aufnahme (Jahre)	17,6
Geschlecht m:w	50:50
durchschnittliche Hilfedauer (Monate)	12,5
Abbruchquote	30%
Anteil Folgehilfen HzE	45%
Anteil positiver Zielerreichung	58,6%
Anteil misslungener Hilfen	10,8%
Positive Entwicklungen in Familie	44%
Positive Entwicklungen beim jungen Menschen	74%
Negativentwicklungen in Familie	40%
Negativentwicklung beim jungen Menschen	17%
Effektstärke bei Familienzielen	keine bis mittel
Effektstärke bei Zielen junger Menschen	keine
Effektstärke bei Integrationszielen	mittel
Kennzahl (nur Abbruch)	Befund
durchschnittliche Hilfedauer	7,2
Anteil misslungener Hilfen	40%
Effektstärke bei Familienzielen	gering-mittel
Effektstärke bei Zielen junger Menschen	negativ
Effektstärke bei Integrationszielen	gering
Kennzahl (ohne Abbruch)	Befund
durchschnittliche Hilfedauer	15,1
Anteil positiver Zielerreichung	75%
Effektstärke bei Familienzielen	mittel bis hoch
Effektstärke bei Zielen junger Menschen	mittel
Effektstärke bei Integrationszielen	hoch

Zusammenfassung und Hinweise:

Anders als bei anderen stationären Hilfen ist die Abbruchquote tolerabel und die durchschnittlich Wirksamkeit gut. In 11% der Fälle kommt es aber während der Hilfe zu Negativentwicklungen der jungen Menschen. In diesem Umfang ist das möglicherweise nicht zu vermeiden. Solche Entwicklungen werden offensichtlich frühzeitig erkannt und führen zu Abbrüchen zu einem angemessen frühen Zeitpunkt. Kein gravierender Handlungsbedarf.

4.9.16 Im Einzelnen: stationäre ISE (Leistungstyp 13)

Wegen geringer Fallzahl ist keine differenzierte Auswertung möglich.

4.10 Wirkungsevaluation bei Clearingfällen (KCD = Krisenmanagement, Clearing, Diagnose)

Die statistischen Auswertungen bezogen sich bisher ausschließlich auf Regelleistungen im Rahmen der Hilfen zur Erziehung. Ausgeschlossen wurden Interventionen, die nicht das Ziel hatten, Probleme lösen zu helfen, sondern die im Vorfeld dazu beitragen sollten, das Problem erst einmal zu verstehen und zu überprüfen, welche Hilfen notwendig und geeignet sind. Oft reichen in Krisensituationen oder in sehr verworrenen Problem-lagen die Möglichkeiten der fallverantwortlichen SozialarbeiterInnen nicht aus, diese Fragen im „laufenden Geschäft“ mit den Klienten zu bearbeiten. Auch muss manchmal unmittelbar eine Gefahr abgewehrt werden, ohne dass bereits genug Informationen vorliegen, um ein tragfähiges Schutzkonzept in Kraft zu setzen.

Die angestrebten Wirkungen von KCD-Maßnahmen sind andere als die normaler Erziehungshilfen. Eine KCD-Intervention ist dann erfolgreich, wenn

- die Probleme am Ende besser verstanden werden (sozialpädagogische Diagnose),
- für eine folgende Hilfeplanung Ressourcen, Ziele, Wünsche der Klienten und tragfähige nächste Schritte klar sind und
- niemand zu Schaden gekommen ist.

Entsprechend ist das Evaluationsinstrument für KCD-Fälle ein anderes. Die fallverantwortliche Fachkraft im Jugendamt muss zu Beginn entscheiden, ob es sich um eine Clearing-Beauftragung handelt, welches die Clearinganlässe und Ziele sind und in welchem Setting die Maßnahme durchgeführt wird.

In dem Projektzeitraum wurden 611 stationäre Hilfen mit dem Ziel eines Clearings begonnen und 465 beendet. Bei ambulanten Clearingfällen gab es 490 Beginn-Einschätzungen und 414 Ende-Einschätzungen. Ob die WIMES-Stichprobe eine repräsentative Stichprobe darstellt und wie viele Clearingfälle es tatsächlich gegeben hat, wissen wir nicht, weil wir keine Vergleichsstatistiken haben.

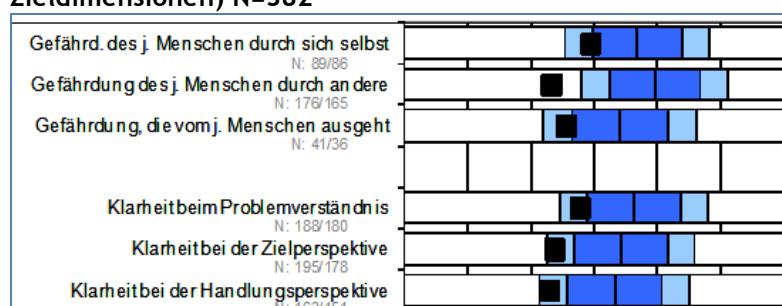
Da die Dauer einer Clearingmaßnahme deutlich unter dem Datenerfassungszeitraum von zwei Jahren liegt, ist es erlaubt und sinnvoll, eine Längsschnittuntersuchung zu machen. Es werden in der Folge also nur solche Maßnahmen erfasst, die in 2011/2012 begonnen und beendet wurden. Nach diesem Kriterium wurden 382 Maßnahmen stationär und 325 ambulant durchgeführt.

Stationäre KCD-Maßnahmen

Die stationären Clearingmaßnahmen dauern durchschnittlich (Median) 2 Monate. 34 % der Clearingmaßnahmen werden vorzeitig abgebrochen. 44 % der Maßnahmen werden während oder im Anschluss an andere Hilfen durchgeführt, offenbar zur Klärung der Ursachen, warum die laufende Hilfe nicht wirkt.

Insgesamt werden die Maßnahmen zu 50 % als erfolgreich eingestuft. In 24 % der Fälle sehen die Fallverantwortlichen im Jugendamt am Ende keinen Nutzen in der Maßnahme. Das zeigt sich auch bei den Effektstärken (Abbildung 35). Am wirksamsten ist die Maßnahme bei der Sicherstellung des Kindeswohls (Abbau von Gefährdungen des jungen Menschen). Der Zugewinn nach Klarheit beim Fallverständen und bei der Ziel- und Handlungsperspektive ist mittel-effektiv. 30 % der Fälle sind weiterhin völlig unklar.

Abbildung 35: Wirksamkeit (Effektstärken) bei stationären Clearing-Maßnahmen (Veränderung bei den Zieldimensionen) N=382



Ambulante KCD-Maßnahmen

Ambulante Maßnahmen dauern im Durchschnitt länger (2,9 Monate im Median) und verlaufen stabiler (Abbruchquote 16 %). Zu 56 % erreichen sie ihre Ziele. In 16 % der Fälle ist man am Ende genauso ratlos wie vorher. Das sind in der Mehrzahl die abgebrochenen Maßnahmen. Die Effektstärken sind höher als bei den stationären Maßnahmen (Abbildung 36).

Abbildung 36: Wirksamkeit (Effektstärken) bei den ambulanten Clearing-Maßnahmen (Veränderungen bei den Zieldimensionen) N=325



Übrigens kann man nicht davon ausgehen, dass ambulante Clearingfälle „leichter“ sind als die stationären. Dazu genügt ein Blick auf die blauen Ausgangsbalken in den beiden Grafiken. Nicht einmal beim Kinderschutz (zweiter Balken) ist bei den stationären Fällen eine wesentlich höhere Gefährdung festzustellen.

Warum es zu einem Clearingauftrag gekommen ist, wie die Entscheidung zwischen ambulant und stationär gefallen ist und warum das Hauptziel der Klärung so wenig beeindruckend erreicht wird, ist aus dem Datenmaterial nicht nachzuvollziehen. Hier wäre eine vertiefende Analyse angebracht.

4.11 Kooperative Wirkungsevaluation von Leistungsträger und Leistungserbringer

Ursprünglich war das Evaluationsprojekt ohne eine Beteiligung der Berliner Träger der Jugendhilfe oder der Liga der Wohlfahrtsverbände konzipiert worden. Allerdings war das Interesse an dem Projekt in der Liga und bei vielen Leistungserbringern sehr hoch. Im Wesentlichen wurde es begrüßt, dass fachliche Themen in Ergänzung zu strukturellen und betriebswirtschaftlichen Themen angepackt wurden. In allen Bezirken sind partnerschaftliche Qualitätsentwicklungsprojekte üblich, so dass bereits im ersten Projektjahr einige Träger als Teilnehmer in dieses Projekt aufgenommen wurden.

Eine gemeinsame Evaluation macht aus mehreren Gründen Sinn:

- Die Validität der Daten steigt durch ein Vier-Augen-Prinzip.
- Ergebnisse werden besser verstanden und leichter akzeptiert, wenn sie auf den eigenen Daten basieren.
- Die Arbeitsbelastungen einer wirkungsorientierten Dokumentation können zwischen Jugendamt und Leistungserbringer aufgeteilt werden.
- Die Hilfeplanung kann durch ein gemeinsames System von Definitionen, Strukturierungshilfen, Maßstäben und Dokumentationsinstrumenten verbessert und erleichtert werden.
- Die Qualitätsentwicklung der Jugendhilfehilfe nach § 79a SGB VIII erfordert eine gemeinsame Erarbeitung von Maßstäben, Instrumenten und Maßnahmen der Sicherstellung definierter Qualität.
- Die Kennzahlen einer Wirkungsevaluation können eine Faktenbasis für die Qualitätsdialoge nach § 78a ff im Berliner Verfahren liefern.

Aus diesen und anderen Gründen wurde beschlossen, das Projekt in seinem dritten Abschnitt (ab 1.10.2012) für Leistungsanbieter zu öffnen, um Erfahrungen mit einer kooperativen Wirkungsevaluation zu machen. 80 Träger schlossen entsprechende Vereinbarungen mit 4 Bezirken¹³, nahmen an Schulungen teil und bereiteten sich technisch auf eine datenschutzgerechte, verschlüsselte Dokumentation vor.

Durch die Schwerpunktverlagerung im Projektverlauf nach dem 17.1.2013 konnte diese Kooperation nicht im geplanten Umfang durchgeführt und erprobt werden.

Datenstand (Stichtag 1.5.2013)

Durch die vorzeitige Beendigung der Datenerfassung in den Bezirken können wir statt auf 5.000 zu erwartende gemeinsam bearbeitete Fälle nur auf einen Datensatz von 1.443 (13,8% der Gesamtstichprobe) aus unsystematisch gemeinsam erhobenen Fällen zugreifen. Trotzdem erscheint uns diese Stichprobe ausreichend und das Thema der Kooperation so wichtig, dass wir einige empirische Tendenzen darstellen wollen. Die Ergebnisse sind mit ausgewählten Einrichtungsvertretern und den Geschäftsführern der Liga diskutiert worden, die sich auch nach den enttäuschenden Projekterfahrungen einstimmig für eine kooperative wirkungsorientierte Steuerung ausgesprochen haben.

Die Wahrscheinlichkeit, kooperativ zu evaluieren

Der Anteil der kooperativen Fälle liegt je nach Hilfeart unterschiedlich hoch. Die unterschiedlichen Kooperationsanteile sind statistisch signifikant ($p < 0.001$, Tabelle 9). Um die Frage zu klären, ob es irgendwelche Fallmerkmale gibt, die einen Einfluss auf die Wahrscheinlichkeiten haben, dass der Fall kooperativ evaluiert wird, kommen wir zu folgenden signifikanten Zusammenhängen:

¹³ Steglitz-Zehlendorf entschied, die Träger zwar zu informieren, eine kooperative Evaluation aber erst zu einem späteren Zeitpunkt in Angriff zu nehmen.

Tabelle 9: Zusammenhang zwischen kooperativer Dokumentation und Hilfeart

Hilfeart	kooperativ dok.
stationär	8,1%
teilstationär	9,2%
ambulant	17,5%
über alle Hilfen	13,8%

Außer der Bezirkszugehörigkeit und der Hilfeart hängt die Kooperationswahrscheinlichkeit signifikant ($p<0,01$) von den Ausgangsbedarfen ab:

- **Stationäre Fälle** werden eher kooperativ dokumentiert, wenn es einen deutlich höheren Bedarf gibt in den Zieldimensionen *Versorgung und Sicherheit in der Familie, Sozialverhalten, Lernen und Leisten, Eigenverantwortung des jungen Menschen*.
- **Teilstationäre Fälle** mit kooperativer Dokumentation weisen einen höheren Eingangsbedarf aus bei *psychischer Stabilität und Lernen und Leisten*.
- **Ambulante Fälle** werden eher kooperativ dokumentiert, wenn es einen hohen Bedarf gibt bei allen Familiendimensionen, Lernen und Leisten, Eigenverantwortlichkeit des jungen Menschen, Teilhabechancen, Rechten und Entwicklungsbedingungen.

Bei einer tiefer gehenden multivariaten Analyse (Klassifizierung mittels Entscheidungsbaum mit der Frage, wann die Kooperationswahrscheinlichkeit am höchsten ist) zeigt sich dieses Bild:

1. Bei den stationären Hilfen ist der Anteil kooperativ dokumentierter Fälle am höchsten bei ganz jungen Kindern und älteren Jugendlichen mit vorheriger Hilfe (27,7 %). Er ist am niedrigsten bei sehr problematischen männlichen Jugendlichen zwischen 12 und 16 Jahren (1,7 %), also genau der Gruppe, bei der die Stabilität und Wirksamkeit der Hilfen am geringsten ist.
2. Bei ambulanten Hilfen ist die Wahrscheinlichkeit einer kooperativen Dokumentation am höchsten, wenn die Problembelastung mittelhoch ist und es sich um einen 8a-Fall handelt oder wenn die Problembelastung mittelhoch ist und vorher eine Hilfe durchgeführt wurde. Gering ist die Wahrscheinlichkeit, wenn es keine Richtungsziele gibt, es weder ein 8a-Fall ist noch vorher eine Hilfe durchgeführt worden ist.

Kooperative Dokumentation und Prozess- und Ergebnisqualität

„Kooperative Dokumentation“ ist ein Zeichen für eine partnerschaftliche Qualitätsverantwortung von Leistungsträger und Leistungserbringer und für ein geteiltes Interesse daran, dass eine Hilfe verlässlich, zielgerichtet und überprüft durchgeführt wird. Ob die gemeinsame Dokumentation tatsächlich zu einem fachlichen Dialog zwischen dem Jugendamt und dem Leistungserbringer geführt hat, ist an den Daten nicht abzulesen. Der gemeinsame fachliche Blick auf den Fall führt im besten Fall dazu, dass

- die Passung der Hilfe zu Beginn und im Verlauf besser ist,
- die Bewertungen realistischer und weniger subjektiv sind,
- Prozessschritte genauer eingehalten werden,
- Frühindikatoren zeitiger beachtet werden.

Wir gehen davon aus, dass im Durchschnitt diese Qualitätsziele durch die kooperative Dokumentation verbessert wurden. Diese Hypothese soll hier überprüft werden, wobei noch einmal darauf hingewiesen werden muss, dass die Aussagen unter dem Vorbehalt stehen, dass die untersuchte Stichprobe nicht repräsentativ ist und die Zeit zu kurz war. Im Folgenden muss die Stichprobe auf die Fälle begrenzt werden, bei denen sowohl am Anfang als auch am Ende gemeinsam bewertet wurde. Ältere laufende Fälle wurden sehr selten einem Leistungserbringer zugeordnet, so dass die Stichprobe sich im Wesentlichen auf kurze Hilfeverläufe beschränkt.

Im Dokumentationsdesign ist vorgesehen, dass nach ungefähr 8 Wochen mit Hilfe der Beobachtungen des Leistungserbringers und den ersten Erfahrungen mit der gewählten Hilfe noch einmal Bedarf und Eignung der Hilfe geprüft werden. Zu diesem Zeitpunkt dürfte auch der Klient eine konkrete Vorstellung von der geplanten Hilfe haben. Die Hilfe wird gegebenenfalls noch einmal angepasst oder so fortgesetzt, wie sie anfangs geplant wurde. Die Richtungsziele müssen spätestens jetzt in konkrete Handlungsziele übersetzt werden und die genaue methodische Durchführung wird mit den nächsten Schritten mit den Verantwortlichkeiten verabredet.

Nicht in allen Fällen wurde dieser Zeitpunkt A2 dokumentiert. Es gibt einen signifikanten Zusammenhang zwischen dem Einbezug der Leistungserbringer und der Wahrscheinlichkeit, dass A2 tatsächlich dokumentiert wurde (Tabelle 10). Unsere Hypothese, dass eine kooperative wirkungsorientierte Dokumentation die Prozessqualität verbessert, wurde bestätigt.

Tabelle 10: Zusammenhang zwischen Kooperation und Überprüfung nach 8 Wochen

Dokumentation	A2 durchgeführt
kooperativ	86,5%
ohne Einbezug des Leistungserbringers	68,1%
insgesamt	71,6%

Kooperation und Effektivität

Die folgenden Auswertungen stellen die Daten, die kooperativ erhoben wurden, in Beziehung zu allen anderen Berliner Daten.

Einen deutlichen Unterschied gibt es bei den Abbruchquoten (Tabelle 11) und der Einschätzung der Zielerreichung im Nachhinein, insbesondere bei den stationären und teilstationären Hilfen. An dieser Stelle kann nicht unterschieden werden, ob die Validität der Angaben durch das Vier-Augen-Urteil besser ist, oder ob tatsächlich die Hilfen stabiler laufen, wenn sie verlässlich kooperativ gesteuert und bewertet werden. Für einen stabilen Verlauf würde sprechen, dass die Hilfedauer steigt (Tabelle 12).

Tabelle 11: Abbruchquote und Kooperation bei der Dokumentation

Hilfeart	Dokumentation koop.	Alle Fälle
stationär	33,5%	42,9%
teilstationär	30,8%	40,7%
ambulant	27,5%	24,6%
insgesamt	28,5%	30,0%

Tabelle 12: Hilfedauer und Kooperation bei der Dokumentation

durchschnittliche Hilfedauer (Median, Monate)	Dokumentation koop.	Alle Fälle
stationär	15,5	12,5
teilstationär	21,8	12,8
ambulant	12,0	12,0

Zusammenfassung und Hinweise:

Es gibt Hinweise darauf, dass eine gemeinsame wirkungsorientierte Dokumentation von Fallverantwortlichen aus dem Jugendamt und Steuerungsverantwortlichen aus den Einrichtungen und Diensten der Leistungserbringer die Prozess- und Ergebnisqualität der Hilfen verbessern. Dieses zeigt sich besonders in einem positiven

Einfluss auf die stationären und teilstationären Abbruchquoten¹⁴. Die Wirkfaktoren dieses Zusammenhangs könnten darin liegen (Hypothesen), dass

- kooperative Fälle differenzierter erfasst werden,
- die Hilfen passender sind,
- die Hilfeüberprüfung und -fortschreibung verbindlicher und fristgerechter durchgeführt wird,
- nicht Erfolg versprechende Hilfen früher beendet oder verändert werden,
- tendenziell Hilfen erfolgreicher verlaufen und deswegen bei den Klienten eine höhere Akzeptanz erzielen.

Es müsste überprüft werden, ob bei stationären Hilfen nicht ein Artefakt dadurch entsteht, dass die kooperativen Fälle eher ortsnahen Hilfen in Berliner Einrichtungen sind und ortsfreien Hilfen, die abbruchriskanter sind, eher nicht kooperativ dokumentiert wurden. Dieses führt zu einer weiteren vagen Hypothese, dass möglicherweise ortsfreie Unterbringungen nicht so genau gesteuert werden (können), und nicht so stabil und erfolgreich verlaufen.

Sollten sich diese Hypothesen erhärten, würde dieses zur Erreichung einer besseren Ergebnisqualität bedeuten, dass eine gemeinsame wirkungsorientierte Steuerung durch Jugendamt und Leistungserbringer im Hilfeplanverfahren zu einem Standard erhoben werden sollte. Für die Besonderheiten einer ortsfreien Unterbringung ließen sich Lösungen finden, indem z.B. der Katalog geeigneter Einrichtungen begrenzt wird und in Vereinbarungen mit den Einrichtungen eine Kooperation bei der wirkungsorientierten Steuerung verbindlich gemacht werden, oder die Einrichtungen zumindest ihre Kennzahlen offen machen.

Zur Qualitätsentwicklung der Angebote sollten auf der Basis der ermittelten Kennzahlen evidenzbasierte Qualitätsdialoge geführt werden. Mangels Fallzahl ist dieses aufgrund der Ergebnisse aus dem vorliegenden Projekt nicht möglich. Es gibt aber einige Berliner Einrichtungen, die schon seit Jahren Ergebniskennzahlen ermitteln, mit denen solche Wirkungsdialoge erprobt werden können. Dadurch würden die Bemühungen von Leistungsträgern, ihre Angebote wirksam gestalten zu wollen, gewürdigt.

¹⁴ Dieses bestätigt den Befund auf der oben erwähnten ABIE-Studie ((Tornow & Ziegler, 2012)), in der sich eine qualitätsbezogene vertrauensvolle Zusammenarbeit zwischen dem Jugendamt und einer stationären Einrichtung als starker Wirkungsfaktor zur Erhöhung der Stabilität und Passung der Hilfen hat nachweisen lassen.

5 Implementierung wirkungsorientierter Kennzahlen - Praxiserfahrungen aus Sicht des e/l/s-Instituts

Die Erfahrungen, die beteiligte Akteure im Laufe des Projektes gemacht haben, dürften aus unterschiedlichen Perspektiven, Interessenlagen und Rollen recht unterschiedlich ausfallen. Hier kann ich nur aus dem eigenen subjektiven Blickwinkel berichten. Das Ziel dieses Abschnitts ist es, zusammenzustellen, was aus dem Prozess für zukünftige Projekte gelernt werden kann.

Das Projekt stieß von Anfang an bei allen Beteiligten auf großes Interesse. Positiv wurde herausgestellt, dass man froh sei, wieder einmal ein fachliches Thema in Angriff zu nehmen. Offensichtlich ist verstanden worden, dass eine systematische Beschäftigung mit der Ergebnisqualität von Erziehungshilfen nicht einem betriebswirtschaftlichen Managerialismus entspringt, sondern dass es um die Sicherung und Weiterentwicklung fachlicher Standards geht, mit dem Ziel, Nutzer und Interessengruppen mit den besten Leistungen zu versorgen, die zurzeit möglich sind.

Arbeitsbelastung in den Jugendämtern

Hinderlich waren für das Projekt die ohnehin hohe Arbeitsbelastung in den Regionalteams der Bezirksjugendämter und die Mehrarbeit, die eine Projektteilnahme mit sich bringt. Offensichtlich ist das Projekt in manchen Bezirken mit der angespannten Personalsituation in Konflikt geraten. Überlastungen, nicht besetzte Stellen, Fehlzeiten durch Krankheit, zusätzliche Aufgaben durch steigendes Fallaufkommen im Kinderschutz führten zur Ablehnung jeglicher Mehrbelastung.

Das Projekt sei zu einem schlechten Zeitpunkt gekommen, haben manche Führungskräfte in den Jugendämtern erklärt. Es ist nicht entscheidbar, ob es einen anderen, besseren Zeitpunkt gegeben hätte, zumal eine hohe Arbeitsbelastung inzwischen ein Dauerzustand in Jugendämtern ist, nicht nur aufgrund objektivierbarer Sachverhalte, sondern zum großen Teil auch als subjektive Grundstimmung. Bei einigen MitarbeiterInnen gab es die Erwartung, dass die Wirkungsevaluation noch im Laufe des Projekts in das Hilfeplanverfahren integriert wird. Letzteres war nicht in der Projektplanung vorgesehen und ließ sich auch nicht so einfach nachlegen. Dazu hätten Datenschutzfragen geklärt, Beteiligungsverfahren eingehalten, IT-technische und Formularanpassungen vorgenommen werden müssen. Auch wären umfangreiche Schulungen nötig gewesen. Dennoch sank die Motivation bei MitarbeiterInnen, weil die erhoffte Arbeitserleichterung auf sich warten ließ.

Ob daraus zu lernen ist, dass zunächst die personellen Ressourcen aufgestockt werden müssten, ehe man ein innovatives Projekt starten kann, bezweifle ich allerdings. Zusätzliche Ressourcen führen nur kurzzeitig zu einer Zufriedenheitszunahme. Wie in der Glücksforschung festgestellt, kalibriert sich ein „beglücktes“ System in kurzer Zeit wieder auf den alten Zustand. Besser wäre es, in Verbindung mit dem Projekt von vornherein Entlastungen zu schaffen, so dass im Erleben von MitarbeiterInnen ein neues Verfahren hilfreich ist, den Arbeitsalltag leichter zu bewältigen.

Einbezug der Leistungserbringer

Bereits in einem frühen Projektstadium signalisierten Leistungserbringer Interesse daran, sich aktiv an der Wirkungsevaluation zu beteiligen und die Datenerhebung zusammen und in Arbeitsteilung mit dem Jugendamt vorzunehmen. Träger versprachen sich einen Vorteil dadurch, dass sie Entwicklungen beim öffentlichen Träger besser mitvollziehen können, dass sich die fachliche Verständigung bei der Aufnahme eines Falles verbessert, dass das Berichtswesen vereinfacht wird und dass die Qualitätsdialoge evidenzbasierter geführt werden. Es gab auch durchaus kritische Haltungen, die aber in der Regel in praktikable und konstruktive Vorschläge einmündeten.

Alle Wohlfahrtsverbände ließen sich inzwischen die Methode wirkungsorientierter Evaluation und Fallsteuerung in Fachveranstaltungen erklären, bewerteten die Berliner oder die Bezirklichen Zwischenergebnisse und äußerten Interesse daran, die Evaluationsergebnisse in Strategien der Qualitätsentwicklung einmünden zu lassen.

In vier Bezirken wurden Arbeitskreise (meistens Unterausschüsse der Arbeitsgemeinschaften nach § 78 SGB) eingerichtet, die eine kooperative Datenerhebung und Bewertung der Evaluation vorbereiteten. Ende 2012 wurden alle Träger, die in den Bezirken Erziehungshilfen durchführen, eingeladen, sich an dem Projekt zu

beteiligen. Achtzig Träger folgten dieser Einladung. Die Zuverlässigkeit der Träger, sich in Schulungen und technischen Vorkehrungen (Verschlüsselung der Daten) darauf vorzubereiten, war beeindruckend. Vielen Skeptikern zum Trotz konnten in nur vier Wochen alle Vorbereitungen abgeschlossen werden. Leider ist es zu dem letzten Teil der Projektphase nicht gekommen. Das Angebot des els-Institutes, in Eigenregie kostenlos in 2013 eine Wirkungsevaluation durchzuführen, nahmen nur einzelne Träger an.

Weiterentwicklung des Verfahren

Auch wenn jede Veränderung der Datenstruktur und der Datenverarbeitung die Vergleichbarkeit über die Zeit (Längsschnittuntersuchungen) und über viele Organisationen hinweg (Benchmarking) gefährdet, muss eine Evaluationsmethode selber anpassungsfähig sein und auf Erfahrungen und Entwicklungen reagieren. Im Verlauf der Berliner Erfahrungen haben sich Veränderungsnotwendigkeiten ergeben, die zu Überarbeitungen in WIMES eingeflossen sind. In mehreren Entwicklungsworkshops wurde mit MitarbeiterInnen der Jugendämter und aus Einrichtungen eine neue Version des WIMES-Verfahrens erarbeitet.

Folgende Veränderungen wurden beschlossen:

- a) Ergänzung von drei familienbezogenen WIMES-Dimensionen, die noch besser den Sozialraumbezug der Hilfen und den Kinderschutz abbilden. Außerdem wurde eine Belastungsdimension eingeführt für die Fälle, bei denen es de facto gar nicht darum geht, neue Kompetenzen zu erwerben, sondern bei denen es ganz basal erst einmal darum geht, die Familie durch eine schwere Belastungssituation zu tragen, ohne dass sie auseinanderfällt oder jemand zu Schaden kommt.
- b) Das Mutter-Kind-Modul wurde vollständig neu aufgebaut, um unterscheidbar zu machen, ob das Kindeswohl gesichert ist, weil es eine fachliche Betreuung gibt, oder weil die Mutter inzwischen an Kompetenzen und Ressourcen gewonnen hat.
- c) Die Abbruchdefinition wurde präzisiert und mit einer Kategorie ergänzt für die Beendigungen, die im Konsens zu einer Veränderung des Hilfesettings führen (Wechsel der Hilfeart oder des Trägers).
- d) Die Kataloge der vorhergehenden und nachfolgenden Hilfen wurden erweitert auf andere Leistungsbereiche (Kinderpsychiatrie, Schule, Arbeitsverwaltung, Eingliederungshilfe). Bei den Jugendhilfeangeboten wurden sozialräumliche Lösungen ergänzt.

Datenschutz

Die Modalitäten der Auftragsdatenverarbeitung wurden nach Intervention des Berliner Beauftragten für Datenschutz und Informationsfreiheit überprüft und führten in der Folge zu erheblichen technischen Anpassungen, um die Anonymität der erfassten Daten zu gewährleisten. Durch die neue Verschlüsselungstechnologie ist nunmehr kein Bezug zu den Klienten mehr möglich.

Die datenschutzrechtlichen Bedenken waren auch in anderen Kommunen aufgekommen und vom ULD in Kiel bestätigt worden. Im gesamten Jahr 2012 wurden technische und rechtliche Anstrengungen unternommen, das Verfahren datenschutzkonform zu machen. Dazu bedurfte es einer ausgefeilten Verschlüsselungstechnologie, deren Entwicklung komplizierter und langwieriger war als zunächst angenommen. Im Oktober 2012 ist es dann gelungen, die technischen und vertraglichen Voraussetzungen für einen ausreichenden Schutz der dokumentierten Daten zu schaffen. In einer konzentrierten Aktion zwischen IT-Fachleuten, Fachbehörden, Einrichtungen und dem IT-Dienstleister des els-Institutes ist es im allerletzten Moment gelungen, den Datenschützern eine Erfolgsmeldung geben zu können.

Inzwischen ist das WIMES-Webportal vom ULD Kiel zertifiziert und trägt das Datenschutz-Gütesiegel, übrigens als erstes und einziges Evaluationsverfahren in der Bundesrepublik. Rückwirkend betrachtet sind wir den Datenschützern dankbar, dass sie auf bestimmte Mängel aufmerksam gemacht haben, weil wir jetzt modernste Verschlüsselungstechnologie anbieten können, die nicht nur die Daten des Jugendamtes sicher macht, sondern auch eine Kooperation mit Partnern ermöglicht. Damit ist es ohne direkte Schnittstellen zwischen zwei IT-Systemen möglich, einen höheren Datenschutz herzustellen als alle anderen bekannten Verfahren und auch als die bisher übliche Kommunikation über Brief, Mail und Fax.

Projektkommunikation

Jeder, der mit Projektmanagement zu tun hat, weiß, wie wichtig es ist, die Projektbeteiligten und auch die Beobachter draußen umfänglich und zeitnah zu informieren. Dieses ist gut gelungen. Folgende Veranstaltungen sind durchgeführt worden:

Art der Veranstaltung	Anzahl
Schulung für MitarbeiterInnen in den Bezirken und bei Leistungserbringern	17
Fachtage / Fachveranstaltungen mit Referat	20
Entwicklungsworkshops	14
Große Koordinationskonferenz (mehr als 4 Teilnehmer)	11
Kleine Koordinationskonferenz (bis zu 4 Teilnehmern)	10

Ohne die Koordinatoren und Koordinatorinnen in den Bezirken und ohne die intensive Begleitung des Projektes in der Senatsverwaltung hätte das Projekt nicht durchgeführt werden können. An dieser Stelle einen großen Dank für die gute Zusammenarbeit und die Unterstützung des Projektes in den jeweiligen Organisationen.

6 Empfehlungen zu einer kontinuierlichen Wirkungsevaluation und einer wirkungsorientierten Steuerung von Hilfen zur Erziehung in Berlin

1. Empfehlung:

Die Senatsverwaltung nutzt die im Projekt gewonnenen Erkenntnisse, um gezielt die Effektivität einzelner Hilfen zu steigern und damit Effizienzgewinne zu erzielen.

Die wichtigsten Hinweise sind in den Zusammenfassungen der empirischen Befunde zu den einzelnen Hilfarten und Leistungstypen gegeben worden (Abschnitte 4.9 ff) und sollen hier nicht im Detail wiederholt werden.

Am steuerungsrelevantesten ist die Senkung der Abbruchquoten bei den stationären und teilstationären Hilfen. Die Zielgruppe, auf die man das Hauptaugenmerk richten sollte, sind junge Menschen im Alter von 12 bis 16 Jahren bei Beginn der Hilfe.

Dafür bieten sich wissenschaftliche Untersuchungen zu den Ursachen und Begleitumständen von Abbrüchen an in Form von Tiefenanalysen anhand von vorhandenen Daten, zusätzliche Datenerhebung, Praxisforschungsprojekte. Des Weiteren sollten Expertengruppen oder „Runde Tische“ mit Leistungserbringern eingerichtet werden.

Bei den teilstationären Hilfen sollte überprüft werden, ob geeignete frühe Hilfen, die schulische Förderungen und eine intensive Elternarbeit mit einbeziehen, nicht Alternativen zur Heimunterbringung sein könnten.

Bei einigen ambulanten Hilfen sollten überprüft werden, warum es trotz guter Wirksamkeit in hohem Maße zu Folgehilfen kommt.

Die Abschlussberichte sollten in der Senatsverwaltung und in den Bezirken dahingehend gelesen und diskutiert werden, welche Verbesserungspotenziale sich daraus ableiten lassen.

2. Empfehlung:

Die Senatsverwaltung und alle Berliner Bezirke führen berlinweit einen einheitlichen Standard zur Messung von Bedarfen der Hilfen zur Erziehung und der Ergebnisqualität von personenbezogenen Leistungen der Jugendhilfe ein.

„Einheitlich“ bedeutet ein gemeinsames Mess- und Dokumentationsverfahren, eine zusammengeführte Datenbank und ein überall gleiches Kennzahlensystem und Berichtswesen. Die Forderung nach Standardisierung bezieht sich nicht auf das methodische Vorgehen im Einzelfall. Letzteres erfordert eine Anpassung an das einmalige Fallverständnis und Raum für ein partizipatives Aushandeln. Die Grundausrichtung der Intervention und ihre Schwerpunkte (Richtung, Ziele) werden allerdings nach einem definierten Standard dokumentiert, ebenso die Zielannäherung oder Zielkorrektur im Verlauf sowie die Zielerreichung am Ende.

Durch die Standardisierung der Daten sind diese vergleichbar und statistisch aggregierbar und auswertbar. Sie bilden die Grundlage einer evidenzbasierten Wirkungsorientierung auf allen Ebenen.

Damit erfüllen die Berliner Bezirke und die Landesjugendamt die Anforderungen des § 79a SGB VIII („Um die Aufgaben der Kinder- und Jugendhilfe nach § 2 zu erfüllen, haben die Träger der öffentlichen Jugendhilfe Grundsätze und Maßstäbe für die Bewertung der Qualität sowie geeignete Maßnahmen zu ihrer Gewährleistung für 1. die Gewährung und Erbringung von Leistungen ... weiterzuentwickeln, anzuwenden und regelmäßig zu überprüfen.“). Darüber hinaus verfügen sie über ein Instrument zur Berlin-weiten Planung der Verteilung von Ressourcen und zu Gestaltung der Jugendhilfeangebote in Berlin.

3. Empfehlung:

Es wird ein Mess- und Dokumentationsverfahren eingesetzt, das einfach zu lernen und ohne großen Aufwand in der Praxis einzusetzen ist.

Man sollte nicht den Anspruch haben, die gesamte Komplexität der sozialen Arbeit abzubilden und auch noch Ursachenforschung im großen Stil zu betreiben. Weniger ist manchmal mehr. Der Dokumentationsaufwand für die Wirkungsdaten sollte pro Fall eine Stunde nicht überschreiten. Dieser zeitliche Aufwand sollte durch Vereinfachung der Gesamtdokumentation und der Kommunikation mit dem Leistungserbringer eingespart werden, so dass der Netto-Zusatzaufwand für die Datenerhebung gleich Null ist.

Der Schulungsaufwand sollte einen halben Tag nicht überschreiten. Das Verfahren muss selbsterklärend und durch Anwendung erlernbar sein.

Die Menge der Informationen sollte auf das absolut Notwendigste beschränkt werden. Es ist inzwischen wissenschaftlich erwiesen, dass ab einer gewissen Datenmenge die Validität wieder sinkt und Entscheidungen schlechter werden (Gigerenzer, 2013)(Gigerenzer & Kober, 2008).

4. Empfehlung:

Die wirkungsorientierte Dokumentation unterstützt neben einer Generierung von strategischen Kennzahlen die Qualitätssicherung im Hilfeplanverfahren. Die wirkungsorientierte Evaluation wird in das Hilfeplanverfahren integriert.

Die wirkungsorientierte Dokumentation ist Teil des Hilfeplanungsprozesses. Die Bedarfseinschätzung vor der Hilfe wird im Zusammenhang mit dem Fallverständnis und dem Hilfeentscheidungsprozess dokumentiert. Die Einschätzung nach 8 Wochen wird von den Leistungserbringern vorgenommen und nach Sichtung durch den Fallverantwortlichen im Jugendamt quittiert oder zur Grundlage eines Hilfeplangespräches gemacht. Das Gleiche gilt für die Entwicklungseinschätzungen im Hilfeverlauf und für die Beendigung der Hilfe. In den Fällen, in denen kein Berliner Träger in die Dokumentation eingebunden ist, dokumentiert der/die Fallverantwortliche nach den Berichten der Leistungsbringer und nach eigener Inaugenscheinnahme.

Durch die Integration in das Hilfeplanverfahren wird der Zusatzaufwand für die Evaluation minimiert, da die entsprechenden Informationen ohnehin im normalen Hilfeplanverfahren erhoben und dokumentiert werden müssen.

Der zusätzliche Nutzwert im Case-Management wird erreicht durch...

- Strukturierung im Fallverständnis,
- Vereinfachung und Vergleichbarkeit im Genehmigungsverfahren,
- einfache Fallübergabe bei Zuständigkeitswechsel, wobei alle Hilfebeteiligten jederzeit die Verantwortlichkeit sehen und automatisch über Änderungen informiert werden,
- einfache Fallübergabe bei Bezirkswechsel und Fortdauer der Hilfe
- Vereinfachung und Sicherstellung der Kommunikation mit Leistungserbringern¹⁵,
- Sicherstellung eines Verlaufscontrollings,
- Strukturierung und Konzentration der Verlaufsberichte,
- Nutzen von Frühindikatoren und Risikoeinschätzungen,
- Schlussbewertung bei Hilfebeendigung (Feedback und Lernen aus Erfahrung).

5. Empfehlung:

Die wirkungsorientierte Dokumentation wird kooperativ mit den Berliner Leistungserbringern vorgenommen.

Diese Empfehlung ergibt sich folgerichtig aus der 2. Empfehlung. Die Schnittstelle zwischen Jugendamt und Leistungserbringer ist für das Gelingen der Hilfen von entscheidender Bedeutung.

¹⁵ Dieses würde Berlin-weit zu einer deutlichen Arbeitsersparnis führen. 4,6% der ambulanten Hilfen enden durch Umzüge der Klienten in andere Bezirke, die den Fall zurzeit neu anlegen müssen. Das entspricht ungefähr 300 Fällen pro Jahr in Berlin.

Die Vorteile einer kooperativen Dokumentation liegen auf der Hand:

- Notwendige Verständigung auf gemeinsame Maßstäbe und abstimmte Ziele
- Hohe Genauigkeit bei der Beauftragung der Leistungserbringer
- Hohe Validität der gewonnenen Daten durch ein Vieraugenprinzip
- Arbeitsersparnis durch Vermeiden einer Paralleldokumentation
- Weniger Informationsverlust an der Schnittstelle Leistungsträger-Leistungserbringer, im Vergleich zu der Anwendung jeweils eigener Dokumentationssysteme.

Mit den Trägern muss vereinbart werden, dass mit den Daten keine Trägerbewertung vorgenommen wird und diese zu einseitigen Veränderungen der Belegungspraxis führen. Die Wirkungsberichte der Leistungsanbieter müssen in den Bezirken partnerschaftlich und offen diskutiert werden, mit dem Ziel einer gemeinsamen Qualitätsentwicklung. Es bietet sich an, dieses innerhalb des Berliner Verfahrens für Qualitätsentwicklungsdialoge zu verankern. Generelle sich ergebende Themen aus der Wirkungsevaluation sollten in den Arbeitsgemeinschaften nach § 78 SGB VIII besprochen werden oder Gegenstand einer Abstimmung mit der Liga der Wohlfahrtsverbände sein.

Es sollte vereinbart werden, wie die Aufwände für die Teilnahme der Leistungserbringer finanziert werden. Die Arbeitszeit für die Träger würden wir als Teil der bereits jetzt vereinbarten Leistungen im Zusammenhang mit der Hilfeplanung und der Qualitätsentwicklung ansehen. Die Kosten für das (technische) Verfahren tragen die Senatsverwaltung und die Bezirke. Außer der Arbeitszeit zu Vorbereitung der technischen Erfordernisse entstehen den Trägern keine zusätzlichen Kosten. Eventuell kommt es durch Arbeitserleichterungen und Wegfall von Kommunikationskosten zu Kostenersparnissen, die dem Träger zugestanden werden sollten.

Bei entsprechenden Vereinbarungen mit Berliner Trägern¹⁶ kann die überwiegende Zahl der Fälle auf diese Weise bearbeitet werden. Lediglich stationäre Unterbringungen außerhalb von Berlin werden allein vom Fallverantwortlichen dokumentiert. Es sollte aber überlegt werden, ob mit stationären Einrichtungen außerhalb Berlins, die häufig aus Berlin belegt werden, Vereinbarungen über eine gemeinsame wirkungsorientierte Steuerung getroffen werden. Dadurch könnte in der Gruppe der Klienten mit dem höchsten Abbruchrisiko ein Effizienzgewinn erzielt werden.

Die datenschutzrechtlichen Vorbereitungen einer solchen kooperativen Dokumentation sind bereits getroffen, indem sichergestellt ist, dass auch seitens der Leistungsanbieter alle sensiblen Daten verschlüsselt sind. Auch die Zuweisung des Falles ausschließlich an die fallzuständige Person seitens des Leistungserbringers im Gegensatz zu der bisherigen Übersendung schriftlicher Informationen an die Einrichtung dient dem Vertrauenschutz der Klienten.

6. Empfehlung

Mittelfristig wird die Wirkungsevaluation auf die Beteiligung der Klienten ausgeweitet. Am Ende der Hilfe¹⁷ findet eine Befragung zur Zufriedenheit mit dem Verlauf und mit dem Nutzwert der Hilfe statt. Die Einschätzungen der Fachkräfte werden den Klienten transparent gemacht. Die Klienten bekommen bei der Hilfenvorbereitung Informationen über die Wirksamkeit der geplanten Hilfen und übernehmen Mitverantwortung für die Wirkung der geplanten Hilfe.

¹⁶ Vier große Anbieter von stationären Jugendhilfeleistungen evaluieren bereits seit mehreren Jahren ihre Wirksamkeit mit dem WIMES-Verfahren. Hier würde also know-how von zwei Seiten zusammenkommen.

¹⁷ Eine Wirkungseinschätzung kann ein Klient erst im Verlauf und am Ende der Hilfe vornehmen. Die Partizipationspraxis zu Beginn der Hilfe ist von unberührt.

Im Sinne der Partizipation der Klienten und der Adressatenorientierung von Hilfeangeboten sollte die Bewertung der Wirkungen und der Verantwortung für das Gelingen einer Intervention mit dem Hilfeadressaten geteilt werden. Dieser Aspekt ist in dem Pilotprojekt bisher zu kurz gekommen, sollte aber im Blick behalten werden.

In den einzelnen Prozessschritten sollten die Einschätzungen durch die Fachkräfte seitens des Jugendamtes und der Leistungserbringer den Klienten erklärt werden. Es gibt Jugendämter, die bereits gute Erfahrungen damit gemacht haben, die WIMES-Dimensionen gemeinsam auszufüllen. Dieses wird nicht in jedem einzelnen Fall möglich sein, sollte aber angestrebter Standard sein.

Am Ende der Hilfe sollte eine schriftliche anonyme Befragung der Klienten (Sicht der Eltern und der jungen Menschen) gemacht werden, deren statistische Auswertung mit den Wirkungsberichten zusammen in die Qualitätsentwicklung einfließt. Für Kinder und Jugendliche werden unterschiedliche Bögen eingesetzt. Die Bögen für Eltern liegen auch in anderen Sprachen vor. Eine mündliche Befragung der Klienten zu ihrer Sicht im letzten Hilfeplangespräch reicht nicht aus, um zu Wirkungskennzahlen aus Kundensicht zu gelangen. Erstens fühlen sich Klienten möglicherweise nicht frei, offen und ehrlich zu antworten und zweitens sind mündliche Antworten nicht valide für Auswertungen zu quantifizieren¹⁸.

Wenn es definierte Risikoeinschätzungen im Verlauf gibt, beispielsweise Abbruchrisiken, sollten diese mit den Klienten offen diskutiert werden. Die Befürchtung, dass man damit Klienten erst auf „dumme Ideen“ bringe, hat sich nicht bewahrheitet. Im Gegenteil: Das offene Gespräch und der Einbezug der Klienten in die Verantwortung für das Gelingen einer Hilfe sind bereits risikomindernd.

7. Empfehlung

Die Daten sollten zeitnah elektronisch erhoben werden und direkt für eine statistische Auswertung zur Verfügung stehen. Das Erhebungsinstrument sollte wechselnden Bedürfnissen leicht angepasst werden können.

Grundsätzlich können Evaluationsdaten auch über Erhebungsbögen (Papier-Bleistift-Verfahren) erhoben werden. Der datenlogistische Aufwand ist dabei allerdings sehr hoch, denn die Informationen müssen zur Datenverarbeitung noch einmal in Datenbanken übertragen werden. Das bedeutet Kosten, Fehlerquellen und zeitliche Verzögerungen.

Nur die zeitnahe (maximal nach 2 Wochen) Dokumentation sichert eine ausreichende Validität der Daten.

Eine direkte elektronische Datenerfassung verringert den Aufwand, gestattet zeitnahe (auch unterjährige) Auswertungen und sichert durch Plausibilitätskontrollen und Fristenerinnerungen die Validität der Daten. Elektronisch erfasste Daten können außerdem leichter in Formulare und andere Datenbanken (z.B. Fachsoftware) eingefügt werden.

Eine prospektive Evaluation, also die Datenerfassung von Anfang der Hilfe an, ist einer ausschließlich retrospektiven Evaluation (im Rückblick vom Ende her) unbedingt vorzuziehen. Rückblickende Bewertungen können zwar ergänzend genutzt werden, als alleinige Daten zu Bewertung von Wirkungen sind sie aber untauglich.

Bei der elektronischen Speicherung von Daten ist unbedingt der gesetzliche Datenschutz einzuhalten. Dieses betrifft die Datensparsamkeit genauso wie die Sicherheit, dass unbefugte Dritte keinen Zugriff auf personenbezogene Daten bekommen. Dazu sind ein differenziertes Rechtekonzept, ein Löschkonzept und die Verschlüsselung aller Namen, Geburtsdaten und Textfelder nötig. Nur die mit dem Fall betrauten verantwortli-

¹⁸ Es sei denn, man nutzt halbstandardisierte Interviewverfahren.

chen Personen sind in der Lage, einen Zugriff auf die Daten vorzunehmen und die Informationen zu entschlüsseln.

8. Empfehlung:

Die Wirkungsevaluation und wirkungsorientierte Steuerung sollte mittels einer speziellen Software webbasiert erfolgen.

Der Einsatz einer Fachsoftware mit integrierter Wirkungsevaluation und Fallsteuerung ist hierfür aus folgenden Gründen weniger geeignet:

- Eine Vernetzung mit den Einrichtungen und Fachkräften aus Einrichtungen und Diensten zur kooperativen wirkungsorientierten Dokumentation und Steuerung ist mit einer Fachsoftware nicht zu leisten. Es gibt zwar technische Möglichkeiten, Schnittstellen zu der Software eines Trägers zu schaffen oder einzelnen Usern VPN-Zugänge zu ermöglichen. Die technischen und datenschutzrechtlichen Probleme bei der Implementierung und Wartung dieser Vernetzung ist aber technisch schwierig und sehr kostenintensiv.
- Prozesse und Verfahren sozialer Arbeit sind sehr dynamisch. Die Anpassungen einer Fachsoftware an Veränderungen, die sich aus der Qualitätsentwicklung von Erziehungshilfen ergeben, sind aufwendig und teuer. Prozesssteuerung in Fachsoftware fest zu verankern bedeutet, Flexibilität und Gestaltungsmöglichkeiten aufzugeben.
- Evaluationsstatistiken müssen zu Steuerungszwecken und für wissenschaftliche Fragestellungen hinreichend komplex sein¹⁹. Vorbereitete Auswertungen in Fachsoftware reichen in der Regel nicht aus. Zusatzprogrammierungen und spätere Erweiterungen, die sich aus der Praxis ergeben, sind kostenaufwändig. Zu empfehlen sind automatisierte erweiterbare Auswertungsroutinen, z.B. in Form von Makros und Befehls-Syntax, die es gestatten, jede gewünschte Auswertung schnell anzufertigen. Bei ausreichendem Datenvolumen sind multivariate Verfahren zu empfehlen, mit denen unter Ausnutzung des gesamten Datenmaterials Prognosen und Strategien berechnet werden können.
- Fachsoftware bietet keine Möglichkeit für ein Benchmarking.

Aus diesen Gründen raten wir zu einem modularen System, bei dem die Funktionen für eine Wirkungsevaluation und die wirkungsorientierte Steuerung mittels einer überschaubaren, flexiblen, webbasierten Anwendung realisiert wird.

Eine Schnittstelle zur Fachsoftware lässt sich herstellen, um

- das Evaluationsprogramm aus der Fachsoftware heraus zu starten und Informationen nicht doppelt erfassen zu müssen,
- die Informationen aus der Anwendung in Formulare der Fachsoftware einlesen zu können,
- die Informationen aus der Anwendung in die Fachsoftware einzuspielen (regelmäßige automatische Synchronisation), um auf den Datenbestand auch offline zugreifen zu können,
- die Informationen aus mehreren Datenbanken verknüpfen zu können (Achtung: Datenschutz),
- Informationen auf sicherem Weg zwischen Leistungsträger und Leistungserbringer auszutauschen.

9. Empfehlung:

Die wirkungsorientierte Evaluation und Fallsteuerung sollte unmittelbar nach dem Projektabschluss (am besten ab 2014) berlinweit fortgesetzt / eingeführt werden.

¹⁹ Das WIMES-Verfahren arbeitet zum Beispiel mit automatisierten Verknüpfungen aus Microsoft ACCESS, EXCEL und POWERPOINT. Für kompliziertere multivariate Auswertungen wird die IBM-Anwendung SPSS eingesetzt. Syntax-Skripte und Makros können Berliner Controllern zur Verfügung gestellt werden. Schulung und Support kann vereinbart werden.

Das vorliegende Projekt hat auch gezeigt, wie schwierig und langwierig die Implementierung eines Evaluations- und Steuerungsverfahrens ist. Es ist mindestens von einem Implementierungszeitraum von 3 bis 5 Jahren auszugehen. Durch das Projekt sind einige der erforderlichen Schritte bereits abgearbeitet:

- MitarbeiterInnen sind geschult und haben eine gewisse Routine in wirkungsorientierter Dokumentation entwickelt.
- Das WIMES-Webportal ist getestet und auf Berliner Anforderungen angepasst worden²⁰.
- Datenschutzfragen sind geklärt.
- IT-Verschlüsselungstechnologie ist eingerichtet (in 4 Bezirken).
- Eine kooperative Evaluation mit Berliner Leistungserbringern ist vertraglich und technisch vorbereitet.
- Die Implementierung in das Berliner Hilfeplanverfahren ist methodisch vorbereitet.
- Führungskräfte haben eine gewisse Erfahrung im Lesen und Bewerten von Wirkungsberichten.

Bei der weiteren Planung sollte bedacht werden, dass die Ergebnisse dieses Projektes nach einer längeren Pause bis zu einer zukünftigen Implementierung einer Wirkungsorientierung wieder verloren gehen und man erneut damit beginnen müsste, Methoden und Verfahrensfahren zu entwickeln, IT- und Datenschutzfragen zu klären, Bezirke und Leistungsträger zu gewinnen.

Eine Beendigung des Projektes ohne konkrete Folgen wäre wie ein Abbruch einer Erziehungshilfe: Es wird von Mal zu Mal schwieriger, Veränderungen zu bewirken.

²⁰ Es fehlen nur noch die Implementierung der Handlungsziele und deren Überprüfung, Formulare und Kopiermöglichkeiten zur Übernahme von Informationen in andere Anwendungen (z.B. Formulare in WORD) und eine Schnittstelle zur Fachsoftware.

7 Anlage Erhebungsinstrument

Das Erhebungsinstrument umfasst Definitionen und Ankerbeispiele für die einzelnen Skalenstufen (Maßstäbe). Eine ausführliche Beschreibung des Verfahrens kann hier nicht gegeben werden. Als Überblick mögen hier der Ankerbogen und das Dokumentationsblatt in Papierform genügen.

1 VERSORGUNG UND SCHUTZ IN DER FAMILIE:

Schutz vor körperlichen Verletzungen und psychischen Traumatisierungen; Grundversorgung mit Essen, Unterkunft, Hygiene, Gesundheit.

Eingeschränkt sicher vor körperlichen und seelischen Verletzungen. Schwacher Schutz durch Vertrauenspersonen. Manchmal Strafen durch Liebesentzug, Nahrungsentzug, Ein- oder Aussperren (Hausarrest, „auf dem Zimmer bleiben“, Fernsehverbot ...).

Eingeschränkte Versorgung. Unregelmäßiges, unausgewogenes Essen und Trinken. Beengte und nicht abgegrenzte Wohnverhältnisse, die möglicherweise nicht immer zugänglich sind. Schlechte Hygiene und medizinische Versorgung.

Mehrfaich in der Woche Schläge und Demütigungen. Lebt ständig in Angst und Unsicherheit. Kein wirksamer Schutz bei Bezugspersonen (zu schwach, keine Bindung).

Vernachlässigung. Der junge Mensch bekommt oft nichts zu essen und zu trinken, allenfalls Geld. Kein ungestörter privater Bereich, unbequemes Bett, Störungen der Nachtruhe. Keine gesicherte Hygiene. Keine gesicherte Gesundheitsversorgung.

Körperlich und seelisch traumatisiert bis zerstört. Brutale Schläge und psychische Gewalt. Kein Schutz bei Bezugspersonen. Auch: wird sexuell missbraucht oder in anderer Weise ausgebeutet.

Unzumutbare Mängel in der Versorgung. Verwahrlosung.

2 ERZIEHUNGSKOMPETENZ DER ELTERN:

Aufsicht, Interesse, Regeln, Konsequenz, Orientierung, Förderung von Selbstständigkeit und Interessen, Freiräume, Erziehung ist dem Alter angemessen.

Die Eltern kümmern sich nicht durchgängig um den jungen Menschen. Der Erziehungsstil ist entweder zu repressiv oder „laissez-faire“. Das Erziehungsverhalten ist oft uneinheitlich und inkonsistent.

Die Eltern kümmern sich wenig um den jungen Menschen. Er wird extrem bestraft (körperlich oder durch psychischen Druck) oder extrem verwöhnt. Er erfährt kaum Orientierung. Er bekommt kaum Gelegenheit oder Unterstützung, Selbstständigkeit und Selbstwert zu entwickeln. Die Eltern setzen kaum Grenzen, auch nicht bei Medien, Alkohol, Nikotin. Sie wissen oft nicht, wo sich der j. M. aufhält und mit wem er Umgang hat. Oder sie sind sehr rigide und lassen wenige Freiräume zur Entwicklung von Eigenständigkeit und Interessen.

Die Eltern verhalten sich ihrem Kind gegenüber gleichgültig und nehmen keinen Einfluss auf seine Entwicklung. Sie setzen keine Grenzen, auch nicht bei Medien, Alkohol, Nikotin, Drogen. Sie wissen meistens nicht, wo sich der j. M. aufhält und mit wem er Umgang hat. Oder sie verhalten sich ihrem Kind gegenüber ausgesprochen rigide, lassen ihm keinerlei Freiräume für eigene Erfahrungen und sanktionieren ihn ständig.

Oder die Eltern pendeln zwischen den Extremen.

3 FAMILIENKOMMUNIKATION UND BEZIEHUNGEN:

Gegenseitiger Respekt, Interesse, Wertschätzung, entspannte Atmosphäre, Generationsgrenzen, durchlässige Außengrenzen, Konfliktlösungen, Hilfe

Die Atmosphäre ist angespannt. Es gibt latente Konflikte, die nicht oder nur indirekt ausgetragen werden. Das Vertrauen untereinander ist eingeschränkt. Grenzen werden einerseits nicht respektiert, andererseits ist die Distanz zwischen Einzelnen groß. Koalitionen. Die Familie hat wenig gute Kontakte nach außen.

Die Beziehungen sind gestört. Es gibt ausgrenzende Koalitionen, die zum Teil schnell wechseln können. Die Atmosphäre ist belastend, der Ton gereizt, anklagend oder depressiv. Konflikte eskalieren schnell und können lange anhalten. Die Grenzen werden einerseits scharf gezogen. Andererseits kommt es zu Einmischungen und Grenzüberschreitungen. Der junge Mensch wird in dem „Beziehungswust“ instrumentalisiert.

Heftige Streitereien sind an der Tagesordnung. Dabei kommt es zu verletzenden Schuldzuschreibungen und zu Gewalttätigkeit der Eltern (gegenüber dem Partner oder dem Kind). Das Klima ist feindselig. Ständig kommt es zu Übergriffen in die Angelegenheiten der anderen. Gefühlskälte und Gleichgültigkeit.

4 PSYCHISCHE UND EMOTIONALE STABILITÄT:

Situationsangemessenes Verhalten, Affekte, Gefühle, Ängste, Grundstimmung, Antrieb, Selbstwert, Selbstkontrolle
 Leichte bis mittlere psychische Beeinträchtigung und Verhaltensauffälligkeiten, die auf Belastung hinweisen. Die Auffälligkeiten beeinträchtigen den jungen Menschen in seinem Wohlbefinden und in seinem Sozial- und Leistungsverhalten.

Deutliche psychische Störungen. Der junge Mensch leidet unter seinen Symptomen und Stimmungen. Diese sind von ihm nicht mehr steuerbar. Die Störungen beeinträchtigen den jungen Menschen in seinem Sozial- und Leistungsverhalten und gefährden die Integration und weitere Entwicklung.

Gravierende psychische Störungen und psychische Erkrankung. Die Störungen behindern den jungen Menschen in seinem Sozial- und Leistungsverhalten und gefährden in hohem Maße seine Integration und weitere Entwicklung.
 Sehr schwache Impulskontrolle, schwere Depression, Affekt- und Denkstörungen, Suizidalität.

5 SOZIALVERHALTEN:

Soziale Regeln, soziales Verständnis, Reflexion von Rolle und Position, Freunde, Kontakte, angemessene Durchsetzung, Impulskontrolle.

Aggressivität, die aber Halt macht, wo andere körperlich verletzt werden könnten. Kleinere Diebstähle oder Unehrlichkeiten (noch nicht delinquent). Eingeschränktes Unrechtsbewusstsein und die Haltung, dass eigene Ansprüche aggressiv durchgesetzt werden dürfen. Wenig Einfühlungsvermögen in andere Menschen. Kein angemessenes Durchsetzungsverhalten.

Nur wenige, unverbindliche und wechselnde Freundschaften. Kein richtiger Freund.

Deutliche Aggressivität und vereinzelt delinquentes Verhalten. Der junge Mensch leitet aus eigenen Bedürfnissen das Recht ab, sich auch gegen die Grenzen und Bedürfnisse anderer durchzusetzen.

Kontakte: Der junge Mensch ist meistens allein. Entweder zieht er sich selbst zurück oder er wird von anderen abgelehnt. - Er pflegt einen Umgang außerhalb der Familie, der die Entwicklung des jungen Menschen gefährdet. - Außenseiter in einer Gruppe, der von den anderen ausgenutzt wird oder Aggressionen auf sich zieht.

Stark aggressives Verhalten, das gezielt oder durch mangelnde Impulskontrolle in Kauf nimmt, andere Menschen zu verletzen. Häufige, gezielte Diebstähle und Raub. Kein soziales Unrechtsbewusstsein. Die Aggressivität kann sich auch gegen nahe stehende Personen richten.

Kontakte: Sozialer Rückzug. - Kontakte außerhalb der Familie, die die Entwicklung des jungen Menschen schädigen (z.B. Schulschwänzen und Streunen, sexuelle Verwahrlosung und Prostitution, Anleiten zu delinquentem Handeln).

6 LERNEN UND LEISTUNG:

Schulbesuch, Ausbildung, Arbeitshaltung, Lernverhalten, Regeln, Hausaufgaben, Lernstand, Konzentration, Motivation, Anstrengungsbereitschaft.

Arbeitsverhalten und Leistungsniveau in Schule und Ausbildung unter den eigenen Möglichkeiten. Schlechte Arbeitshaltung. Bis zu 10 unentschuldigte Fehltage im Schuljahr. Disziplinarische Probleme (Einträge, Klassenkonferenz). Fehlen am Arbeitsplatz. Mehrere Ermahnungen. Mängel bei der Arbeit. Androhung des Rausschmisses. Der junge Mensch setzt sich mit eigenem Leistungsanspruch so sehr unter Druck, dass es zu Lern- und Leistungsstörungen kommt.

Auffälliges schulisches Verhalten. Realisiertes Leistungsniveau weit unter den eigenen Möglichkeiten. Über 10 Tage unentschuldigte Fehltage im Schuljahr. Mehrfach disziplinarische Strafen. (Drohender) Schulwechsel wegen disziplinarischer Probleme. Häufiges Fehlen am Arbeitsplatz. Schlechte Arbeitsleistung. Geringe Motivation, sich auf Regeln am Arbeitsplatz einzulassen. Mehr als einmal aufgrund des Verhaltens an einem Arbeitsplatz rausgeschmissen.

Schulverweigern. Geht allenfalls sporadisch zu Schule und hält dann den schulischen Rahmen nicht ein. Leistungsstand weit hinter seinen intellektuellen Möglichkeiten. Lässt sich auf keine Ausbildungs- oder Arbeitssituation ein.

7 EIGENVERANTWORTUNG:

Verantwortungsübernahme, realistische Selbstwahrnehmung, Lösungen suchen, Zukunftspläne, Freizeit, Medien, Selbständigkeit, Kooperation, Durchhaltevermögen.

Einigermaßen realistische Selbst- und Situationseinschätzung, aber Herunterspielen der Probleme und der eigenen Verantwortung . Schiebt oft Verantwortung ab.

Zu unselbstständig. Der junge Mensch traut sich wenig zu. Gibt schnell auf. Misserfolgsorientiert.

Gefährdet seine Gesundheit. Schlechte Ernährung. Gelegentlich Alkohol, Nikotin.

Kann sich schlecht selbst beschäftigen. Kaum Ideen. Keine richtigen Hobbys, keine Ideen. Im Wesentlichen passiver Medienkonsum.

Kein realistisches Bild der Probleme. Verlässt sich bei Problemen auf andere. In seiner Alltagskompetenz um Jahre zurück.

Schädigt langfristig seine Gesundheit. Ungesunde Ernährung. Regelmäßig Alkohol, Nikotin, Drogen. Bewegungsman gel. Kümmert sich nicht um Hygiene und Gesundheitsvorsorge.

Kann sich sehr schlecht in seiner Freizeit beschäftigen. Keine Hobbys, keine Ideen. Viel passiver Medienkonsum (mehr als 3 Stunden Fernsehen) oder elektronische Spiele mit Suchtcharakter.

Sieht keine Probleme oder leugnet sie. Arbeitet nicht an Lösungen. Macht sich keine Gedanken um die Zukunft. Wehrt sich aktiv gegen Veränderungen.

Schädigt seine Gesundheit. Schlechte Ernährung mit Mangelerscheinungen. Übergewicht, Haltungsschäden. Starker Konsum von Alkohol, Nikotin, Drogen. Fehlende Hygiene. Keine Gesundheitsversorgung.

Völlig unselbstständig. Wehrt sich aktiv, Dinge selbst zu erledigen. Empfindet alles als „Zumutung“, reagiert mit Verweigerung oder Aggression.

Kann überhaupt nichts mit seiner Freizeit anfangen. Exzessiver Medienkonsum.

8 Möglichkeiten zur Teilhabe, Bildung und Freizeitgestaltung:

Bildungsmöglichkeiten, Lerngelegenheiten, Lernanreize, Entfaltungsmöglichkeiten sowie Teilhabe am sozialen und gesellschaftlichen Leben sind sicher gestellt.

Eingeschränkte Möglichkeiten und Unterstützung zur persönlichen, mentalen und kulturellen Entwicklung. Umgebung bietet nur wenig Anregungen und Förderungsmöglichkeiten.

Die sozial-kulturelle Entwicklung wird erheblich beeinträchtigt. Anregungsarme Umgebung. Interessen und Begabungen werden kaum gefördert. Werteentwicklung kaum möglich.

Die soziale und kulturelle Entwicklung des jungen Menschen ist durch das Umfeld umfassend behindert und geschädigt. Keine Gelegenheit und Zugang zu Bildungs- und Freizeitmöglichkeiten. Bildungsfeindliche Umgebung. Im Umfeld werden Werte und andere Kulturen diffamiert.

(Für die 9. Dimension gibt es keine Ankerbeschreibungen)

Beispielansichten aus dem WIMES-Webportal

Abbildung 37: Basisdaten zur Fallerfassung

The screenshot shows a web-based form titled "Basisdaten (Klient und Hilfe)". At the top, there is a navigation bar with links for "Logout", "Start", "Meine Nutzerdaten", "Meine Fälle", and "Partner". The main form contains the following fields:

Vorname:	Anita	Nachname:	Testperson	
Geburtstag:	1998-02-01	Geschlecht:	weiblich	
Krisenmanagement, Clearing, Diagnose:	Nein	inkl. MuKi Modul?:	<input type="checkbox"/>	
Aufnahmedatum	21.11.2010 <input type="button" value="..."/> *			
Hilfeform	<input type="radio"/> ambulant	<input checked="" type="radio"/> teilstationär	<input type="radio"/> stationär	
Hilfeintensität	<input type="radio"/> hoch	<input checked="" type="radio"/> regel	<input type="radio"/> gering	
Spezifische Leistung	<input checked="" type="checkbox"/> mit Schule Beruf	<input type="checkbox"/> mit Therapie	<input type="checkbox"/> mit Familienarbeit	
HzE vorher (Anschluss)	<input checked="" type="radio"/> keine	<input type="radio"/> Stationär	<input type="radio"/> Teilstationär	<input type="radio"/> Ambulant
Intensität im Vergleich zur jetzigen Hilfe	<input type="radio"/> weniger intensiv	<input type="radio"/> genau so intensiv	<input type="radio"/> intensiver	
Spezifische Auswertung T.1	0			
Spezifische Auswertung I.1	<input type="text"/>			

To the right of the form, there is a legend for the "Erklärung" (Explanation) of the service forms:

- Laufender Altfall (ohne Eingabe von Bedarf und Diagnose - die Prozessschritte werden ohne Eingabe gesperrt)
- hoch: stationär Pädagogen pro Platz besser als 1 zu 1,7 ; ambulant mehr als 10 Stunden in der Woche
- regel: stationär Pädagogen pro Platz = 1 zu 1,71 bis 2,3 ; ambulant 5 bis 10 Stunden pro Woche*
- gering: stationär Pädagogen pro Platz weniger als 1 zu 2,3 ; ambulant weniger als 5 Stunden pro Woche

At the bottom of the form are buttons for "Speichern und Weiter", "Abbrechen", and "Löschen".

Abbildung 38: Fallübersicht und Hauptnavigationsansicht

The screenshot shows the "Meine Fälle" (My Cases) overview page. At the top, there are links for "Alle Fälle", "Nicht abgeschlossene Fälle", and "Abgeschlossene Fälle". Below this is a search bar with "Suche:" and "Suchen" buttons, along with a "Filter löschen" button.

On the left, there is a button for "Neuen Fall anlegen" (Create New Case). The main area is a table displaying case details:

	Fallcode	Vorname	Nachname	Verantwortlich	Organisation	Erstellt	Aufnahme	A1	A2	E	Aktion
001	TSTF-JLXX-9905	Julia	Testfall	TestFVLerb2	JugendamtTest	25.11.2010	03.02.2010	1	0	0	
002	NCHN-VRNM-0029	Vorname	Nachname	TestFVLerb2		18.10.2010	17.10.2010	3	3 fertig	0	

Abbildung 39: WIMES-Dimensionen zur Bedarfseinschätzung (hier am Beispiel für Mutter und Kind)

WIMES Falldokumentation (inkl. MK) für: TSTP-NTXX-8801 Anita Testperson weiblich 01.02.1988

Noch 6 Pflichtfelder offen

Anmelddaten / Berdarseinschätzung zur Hilfe / Bedarf und Ziele nach 6 Wochen / Verlauf / am Ende

WIMES / Muki Dimensionen MuKi Hauptauftrag

Wimes Dimensionen (beziehen sich auf Mutter/Vater und deren Herkunfts familie)	Muki Dimensionen (elterliche Funktionen und Entwicklung des Kindes)
Ausgefüllt Am <input type="text"/> <input type="button" value="..."/> *	Ausgefüllt Am <input type="text"/> <input type="button" value="..."/> *
1 Versorgung und Schutz in der Familie 	1 M/V: Beaufsichtigen und vor Gefahren/Gewalt schützen
2 Erziehungskompetenz der Eltern 	2 M/V: Die materielle Lebengrundlage sicherstellen
3 Familienkommunikation und Beziehungen 	3 M/V: Feinfühlig auf die Bedürfnisse des Kindes reagieren
4 Psychische und Emotionale Stabilität 	4 M/V: Den Alltag auf das Kind abstimmen
5 Sozialverhalten 	5 M/V: Die Beziehung zu Dritten für das Kind förderlichen gestalten
6 Lernen und Leistung 	6 Kind: Allgemeiner Gesundheitszustand
7 Eigenverantwortung 	7 Kind: Visuo-motorische Entwicklung
8 Möglichkeiten zur Teilhabe, Bildung und Freizeitgestaltung 	8 Kind: Sozial-emotionale Entwicklung
Rechte: Lebens- und Entwicklungsbedingungen 	9 Kind: Kognitiv-mentale Entwicklung

Speichern **Speichern und weiter**

8 Anlage Datengrundlage

Vergleich der Fallzahlen zwischen dem tatsächlichen Fallaufkommen (ermittelt über ProJug) und den evaluierten Fällen in der Evaluationsdatenbank (WIMES).

Da die Anteile der erfassten Evaluationsdaten nach den Leistungstypen variieren, werden einige Gesamtauswertungen mit gewichteten WIMES-Daten vorgenommen. Dadurch werden Stichprobenverzerrungen ausgeglichen.

A) Anteil der dokumentierten Fälle an allen begonnenen und beendeten Fällen

	begonnene Hilfen			beendete Hilfen		
	WIMES-Ist	Pro-Jug (Soll)	Datenerfassung	WIMES-Ist	Pro-Jug (Soll)	Datenerfassung
1 = ambulante Psychotherapie	402	697	58%	297	644	46%
2 = integrative Lerntherapie	447	801	56%	253	463	55%
3 = ambulante Familientherapie	208	313	66%	170	250	68%
4 = Soziale Gruppenarbeit	315	502	63%	238	472	50%
5 = Betreuungshilfe / Erz.-Beist.	815	1313	62%	664	1164	57%
6 = SPFH	1826	2841	64%	1419	2554	56%
7 = ambulante ISE	39	45	87%	30	48	63%
8 = teilstationär	278	604	46%	213	487	44%
9 = stationär (Schichtdienst)	1075	2287	47%	803	2005	40%
10 = stationär (WAB)	330	354	93%	205	290	71%
11 = stationär (familienanalog)	293	298	98%	169	272	62%
12 = stationär (indiv. / BEW)	268	536	50%	237	429	55%
13 = stationäre ISE	45	30	150% ²¹	46	46	100%
15 = Sonstige ambulante	229	636	21%	176	394	23%
14 = Sonstige stationär	132			91		
16= Sonstige teilstationär	89			70		
Alle Hilfen	6791	11257	60%	5081	9518	53%

Bei 14 bis 16 war keine Zuordnung zu Leistungstypen möglich

²¹ Werte über 100% sind theoretisch nicht möglich. Hier sind Hilfen in WIMES häufiger als in der Dokumentation in ProJug einem bestimmten Leistungstyp zugeordnet worden.

9 Anlage Liste der Auswertungen

Die CD-ROM mit den statistischen Auswertungen, aggregiert über alle fünf teilnehmenden Bezirke, umfasst folgende Powerpoint-Dateien. Lesehilfen für die einzelnen Powerpoint-Dateien finden Sie auf der CD in der Datei [Erläuterungen zum Wirkungsbericht.pdf](#).

Die Dateinamen sind folgendermaßen aufgebaut:

<Datum der Auswertung> <20010 = Berlin> <Teilstichprobe> <Zeitraum ><Filter>.pptx

Tabelle 13: Liste der Auswertung der Gesamt-Daten (Inhalt der Auswertungs-CD)

Dateibezeichnung	Inhalt
▫ 2013 05 07 20010 Berlin ambulante sozäh 2011-12 nur Abbrüche.pptx	Zusammengefasst soziale Gruppenarbeit, Erziehungsbeistandschaft SBFH, ambulante ISE
▫ 2013 05 07 20010 Berlin ambulante sozäh 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin ambulante sozäh 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin ambulante ther 2011-12.pptx	Zusammengefasst ambulante Psychotherapie, Lerntherapie, Familientherapie
▫ 2013 05 07 20010 Berlin ambulante ther 2011-12 nur Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin ambulante ther 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin einzelfallbezogene 2011-12 nur Abbrüche.pptx	alle, außer SPFH und Familientherapie
▫ 2013 05 07 20010 Berlin einzelfallbezogene 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin einzelfallbezogene 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin familienbezogene 2011-12 nur Abbrüche.pptx	SPFH und Familientherapie
▫ 2013 05 07 20010 Berlin familienbezogene 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin familienbezogene 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 Längsschnitt nur Abbrüche.pptx	Im Zeitraum begonnen und abgeschlossen
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 Längsschnitt ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 Längsschnitt.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 mit Bench.pptx	alle, im Vergleich zur Benchmarkstichprobe
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 nur Abbr mit Bench.pptx	alle Abbrüche, im Vergleich zur Benchmarkstichprobe
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 nur Abbrüche.pptx	Alle beendeten Hilfen in dem Zeitraum, egal, wann sie begonnen sind
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 ohne Abbr mit Bench.pptx	alle nicht abgebrochenen, im Vergleich zur Benchmarkstichprobe
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt 2011-2012 ohne Abbrüche_Kopie.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt 2011-2012.pptx	dito
▫ 2013 05 07 20010 Berlin gesamt Abbruch in 3 Mon.	Hilfen, die in dem Zeitraum begonnen und be-

Dateibezeichnung	Inhalt
2011-2012 Längsschnitt.pptx	endet haben, Spezialauswahl. Es erfolgte ein Abbruch innerhalb der ersten 3 Monate.
▫ 2013 05 07 20010 Berlin gesamt Abbruch in 3 Mon. 2011-2012.pptx	Alle Hilfen, die in dem Zeitraum in den ersten drei Monaten abgebrochen wurden
▫ 2013 05 07 20010 Berlin gesamt Abbruch nach 3 Mon. 2011-2012 Längsschnitt.pptx	Alle Hilfen, die in dem Zeitraum nach den ersten drei Monaten abgebrochen wurden
▫ 2013 05 07 20010 Berlin gesamt Abbruch nach 3 Mon. 2011-2012.pptx	dito
▫ 2013 05 07 20010 Berlin KCD 2011-2012 nur Abbrüche.pptx	Krisenintervention -, Clearing -, Diagnostik - Aufträge. Hilfen, die in dem Zeitraum beendet worden sind.
▫ 2013 05 07 20010 Berlin KCD 2011-2012 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin KCD 2011-2012.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-1 amb. Psycho 2011-12 nur Abbrüche.pptx	Leistungsgruppe 1: Ambulante Psychotherapie.
▫ 2013 05 07 20010 Berlin Zus 01-1 amb. Psycho 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-1 amb. Psycho 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-10 st. WG 2011-12 nur Abbrüche.pptx	Leistungsgruppe 10: stationär (Gruppenangebot, Wohngemeinschaft)
▫ 2013 05 07 20010 Berlin Zus 01-10 st. WG 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-10 st. WG 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-10 st. WG ab 12 J. 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-10 st. WG u. 12 J. 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-11 st. fam.analog 2011-12 nur Abbrüche.pptx	Leistungsgruppe 11: familienanaloge Hilfen
▫ 2013 05 07 20010 Berlin Zus 01-11 st. fam.analog 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-11 st. fam.analog 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-11 st. fam.analog ab 12 J. 2011-12.pptx	dito; nur Fälle, die im Alter von über 12 Jahren aufgenommen wurden.
▫ 2013 05 07 20010 Berlin Zus 01-11 st. fam.analog u. 12 J. 2011-12.pptx	Leistungsgruppe 11: Familienanaloge Hilfen
▫ 2013 05 07 20010 Berlin Zus 01-12 st. BEW 2011-12 nur Abbrüche.pptx	Leistungsgruppe 12: Betreutes Einzelwohnen
▫ 2013 05 07 20010 Berlin Zus 01-12 st. BEW 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-12 st. BEW 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-12 st. BEW ab 12 J. 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-2 int. Lernther 2011-12 nur Abbrüche.pptx	Leistungsgruppe 2: Integrative Lerntherapie
▫ 2013 05 07 20010 Berlin Zus 01-2 int. Lernther 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-2 int. Lernther 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-3 amb Fam.ther 2011-12 nur Abbrüche.pptx	Leistungsgruppe 3: ambulante Familientherapie

Dateibezeichnung	Inhalt
▫ 2013 05 07 20010 Berlin Zus 01-3 amb Fam.ther 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-3 amb Fam.ther 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-4 soz Gr.arb 2011-12 nur Abbrüche.pptx	Leistungsgruppe 4: Soziale Gruppenarbeit
▫ 2013 05 07 20010 Berlin Zus 01-4 soz Gr.arb 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-4 soz Gr.arb 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-5 Bet.hilfe 2011-12 nur Abbrüche.pptx	Leistungsgruppe 5: Betreuungshilfe/ Erziehungsbeistand
▫ 2013 05 07 20010 Berlin Zus 01-5 Bet.hilfe 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-5 Bet.hilfe 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-6 soz. Fam.hilfe 2011-12 nur Abbrüche.pptx	Leistungsgruppe 6: sozialpädagogische Familienhilfe
▫ 2013 05 07 20010 Berlin Zus 01-6 soz. Fam.hilfe 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-6 soz. Fam.hilfe 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-7 ISE 2011-12 nur Abbrüche - zu wenig.pptx	Leistungsgruppe 7: ambulante ISE, nicht auswertbar: zu wenig Fälle
▫ 2013 05 07 20010 Berlin Zus 01-7 ISE 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-7 ISE 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-8 Tagesgr 2011-12 nur Abbrüche.pptx	Leistungsgruppe 8: Tagesgruppe
▫ 2013 05 07 20010 Berlin Zus 01-8 Tagesgr 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-8 Tagesgr 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-9 st. Gr.angeb 2011-12 nur Abbrüche.pptx	Leistungsgruppe 10: stationäres Gruppenangebot
▫ 2013 05 07 20010 Berlin Zus 01-9 st. Gr.angeb 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-9 st. Gr.angeb 2011-12.pptx	dito
▫ 2013 05 07 20010 Berlin Zus 01-9 st. Gr.angeb ab 12 J. 2011-12.pptx	dito; Aufnahmealter über 12 Jahre
▫ 2013 05 07 20010 Berlin Zus 01-9 st. Gr.angeb u. 12 J. 2011-12.pptx	dito; Aufnahmealter unter 12 Jahre
▫ 2013 05 12 20010 Berlin gesamt KCD gebunden 2011 2012.pptx	KCD
▫ 2013 05 13 20010 Berlin gesamt KCD gebunden 2011-12 nur Abbrüche.pptx	KCD
▫ 2013 05 13 20010 Berlin Zus 01-8 Tagesgr. 2011.pptx	Leistungsgruppe 8: Tagesgruppe
▫ 2013 05 13 20010 Berlin Zus 01-8 Tagesgr. 2012.pptx	dito
▫ 2013 05 13 20010 Berlin Zus 01-8 Tagesgr. ab 12 J. 2012.pptx	dito; Aufnahmealter über 12 Jahre
▫ 2013 05 13 20010 Berlin Zus 01-8 Tagesgr. u. 12 J. 2012.pptx	dito; Aufnahme alter unter 12 Jahre
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. ohne § 32 2011-12 nur Abbrüche.pptx	dito; in diesen Fällen würde nicht teilstationär angegeben

Dateibezeichnung	Inhalt
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. ohne § 32 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. ohne § 32 2011-12.pptx	dito
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. §32 2011-12 nur Abbrüche.pptx	dito; in diesen Fällen wurde „teilstationär“ angegeben
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. §32 2011-12 ohne Abbrüche.pptx	dito
▫ 2013 05 21 20010 Berlin Zus 01-8 Tagesgr. §32 2011-12.pptx	dito
▫ 2013 05 27 20010 Berlin ges. Alter ab 12 J. 2011-2012.pptx	alle Hilfen ab 12 Jahre Aufnahmealter
▫ 2013 05 27 20010 Berlin ges. Alter u. 12 J. 2011-2012.pptx	alle Hilfen unter 12 Jahren Aufnahmealter
▫ 2013 06 10 20010 mit koop Lerb 2011 2012 mit Berliner Bench.pptx	Alle Berliner Fälle, die zusammen mit Leistungserbringern evaluiert wurden im Vergleich zu allen Hilfen in Berlin
▫ 2013 06 10 20010 mit koop Lerb 2011 2012 nur Abbrüche.pptx	dito
▫ 2013 06 10 20010 mit koop Lerb 2011 2012 ohne Abbrüche.pptx	dito
▫ 2013 06 10 20010 mit koop Lerb 2011 2012.pptx	dito

10 Anlage Abbildungen und Tabellen (im Text)

Abbildung 1: Dokumentationsablauf	16
Abbildung 2: Bedarfsdimensionen nach WIMES.....	17
Abbildung 3: Evaluationsdesign	18
Abbildung 4: Verteilung der Variable „Hauptverursacher von Abbrüchen“ bei ambulanten Hilfen	20
Abbildung 5: Verteilung der Variable „Dauer ambulanter Hilfen“	21
Abbildung 6: Verteilung der Einschätzungen der Entwicklungen während der ambulanten Hilfen.....	21
Abbildung 7: Handlungsrelevanzmatrix und ihre strategische Bedeutung	24
Abbildung 8: Handlungsrelevanzmatrix bei teilstationären Hilfen im Berliner Projekt	24
Abbildung 9: Handlungsrelevanzmatrix teilstationärer Hilfen im übrigen Bundesgebiet	25
Abbildung 10 Effekte in den einzelnen Zieldimensionen; alle Hilfen, alle Daten	28
Abbildung 11: Effekte bei den Richtungszielen ambulanter Psychotherapie	29
Abbildung 12: Effektstärken bei individuell betreutem Wohnen; alle 5 Bezirke; Richtungsziele	29
Abbildung 13: Retrospektive Bewertung der Zielerreichung durch die Hilfe; alle Hilfen, alle 5 Bezirke.....	30
Abbildung 14: Retrospektive Beurteilung der Entwicklungen während der Hilfe (alle Hilfen)	30
Abbildung 15: Nachfolgende Hilfen bei allen beendeten stationären Hilfen	32
Abbildung 16: Nachfolgende Hilfen bei allen beendeten teilstationären Hilfen	32
Abbildung 17: Nachfolgende Hilfen bei allen beendeten ambulanten Hilfen.....	33
Abbildung 18: Zielerreichung von Hilfen in Tagesgruppen	46
Abbildung 19: Effektstärken in den Zieldimensionen teilstationärer Hilfen in den fünf Berliner Bezirken	46
Abbildung 20: Effektstärken in den Zieldimensionen teilstationärer Hilfen in der nicht-Berliner Vergleichsstichprobe.....	47
Abbildung 21: Effektstärken bei den Zieldimensionen nicht-abgebrochenen teilstationären Hilfen in Tagesgruppen in Berlin	47
Abbildung 22: Abbruchrisiko in Berliner Tagesgruppen in Abhängigkeit vom Aufnahmealter (N = 185)	48
Abbildung 23: Abbruchrisiko in nicht-Berliner Tagesgruppen in Abhängigkeit vom Aufnahmealter (N = 2719).48	48
Abbildung 24: Unterschiede in den anfänglichen Problembelastungen bei Tagesgruppenklientel, differenziert nach Berlin und nicht-Berlin	49
Abbildung 25: Bedarfsprofile am Anfang der teilstationären Hilfen	49
Abbildung 26:Anteil der als für die teilstationäre Hilfen relevant eingestuften Dimensionen; Vergleich zwischen Berlin und nicht-Berlin.....	50
Abbildung 27: Anzahl (%) der Zieldimensionen teilstationärer Hilfen im Vergleich zwischen Berlin und Nicht-Berlin.....	50
Abbildung 28: Erzieherische Bedarfe bei stationären Hilfen zur Erziehung.....	53

Abbildung 29: Hilfedauer bei nicht abgebrochenen stationären Hilfen	54
Abbildung 30: Effektstärken bei stationären nicht abgebrochenen Hilfen (nur Richtungsziele).....	55
Abbildung 31: Retrospektive Einschätzung der Entwicklung während nicht abgebrochener stationärer Hilfen	55
Abbildung 32: Zielerreichung bei stationären nicht abgebrochenen Hilfen.....	55
Abbildung 33: Zielerreichung bei stationären abgebrochenen Hilfen	56
Abbildung 34: stationäre Hilfen: Wer bricht ab?.....	56
Abbildung 35: Wirksamkeit (Effektstärken) bei stationären Clearing-Maßnahmen (Veränderung bei den Zieldimensionen) N=382.....	66
Abbildung 36: Wirksamkeit (Effektstärken) bei den ambulanten Clearing-Maßnahmen (Veränderungen bei den Zieldimensionen) N=325.....	67
Abbildung 37: Basisdaten zur Fallerfassung	84
Abbildung 38: Fallübersicht und Hauptnavigationsansicht	84
Abbildung 39: WIMES-Dimensionen zur Bedarfseinschätzung (hier am Beispiel für Mutter und Kind)	85
Tabelle 1: Test eines Zusammenhangs über Mittelwertvergleich.....	22
Tabelle 2: Erklärung statistischer Begriffe	25
Tabelle 3: Effektstärke bei der allgemeinen Bedarfsreduzierung.....	27
Tabelle 4: Ausgangsbedarfe und Effektstärke nach Leistungstypen; alle 5 Bezirke; Richtungsziele	28
Tabelle 5: Abbruchquote im Berliner Projekt (nicht gewichtete Stichprobe)	31
Tabelle 6: Effektivität der Hilfesettings je nach Art der Beendigung	31
Tabelle 7: Abbruchquote und Aufnahmealter bei stationären Hilfen.....	57
Tabelle 8: Hilfedauer in Jahren (Median)	57
Tabelle 9: Zusammenhang zwischen kooperativer Dokumentation und Hilfeart.....	69
Tabelle 10: Zusammenhang zwischen Kooperation und Überprüfung nach 8 Wochen	70
Tabelle 11: Abbruchquote und Kooperation bei der Dokumentation	70
Tabelle 12: Hilfedauer und Kooperation bei der Dokumentation	70
Tabelle 13: Liste der Auswertung der Gesamt-Daten (Inhalt der Auswertungs-CD)	87

11 Anlage Literaturverzeichnis

- Beckmann, C. (2004). *Qualität in der sozialen Arbeit: Zwischen Nutzerinteresse und Kostenkontrolle* (1. Aufl.). Wiesbaden: VS, Verl. für Sozialwiss.
- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (3., überarb.). Berlin [u.a.]: Springer.
- Breyer, F., & Zweifel, P. (1992). *Gesundheitsökonomie*. Heidelberg: Springer-Verlag Berlin.
- Bruhn, M. (1999). *Kosten und Nutzen des Qualitätsmanagements*: Grundlagen, Methoden, Fallbeispiele. München: Hanser.
- Bruhn, M. (2003). *Qualitätsmanagement für Dienstleistungen*: Grundlagen, Konzepte, Methoden (4. Aufl.). Berlin [u.a.]: Springer.
- Dukek, C., & Burmeister, J. (2012). *Qualitätsmanagement im Jugendamt*: ein Prozessmodell für den ASD unter besonderer Berücksichtigung des Datenschutzes. Berlin: Eigenverlag des Deutschen Vereins für öffentliche und private Fürsorge e.V.
- Eversheim, W. (1997). *Qualitätsmanagement für Dienstleister*: Grundlagen, Selbstanalyse, Umsetzungshilfen. Berlin; New York: Springer.
- Gigerenzer, G. (2008). *Das Einmaleins der Skepsis: Über den richtigen Umgang mit Zahlen und Risiken* (4. Aufl.). Berlin: Berliner Taschenbuch-Verl.
- Gigerenzer, G. (2011). *Heuristics*: the foundations of adaptive behavior. Oxford;;New York: Oxford University Press.
- Gigerenzer, G. (2013). *Risiko: wie man die richtigen Entscheidungen trifft*. München: Bertelsmann.
- Gigerenzer, G., & Kober, H. (2008). *Bauchentscheidungen: die Intelligenz des Unbewussten und die Macht der Intuition*. München: Goldmann.
- Jessen, D., Arlt, M., Kurz-Adam, M., Schrappner, C., Landes, B., & Deutsches Institut für Urbanistik Arbeitsgruppe Fachtagungen Jugendhilfe. (2011). *Wer steuert die Hilfen zur Erziehung? Die Politik, der ASD, die Interessen der freien Träger, der Kostendruck, die Wirklichkeit ...? Dokumentation der Fachtagung am 7. und 8. April 2011 in Berlin*. Berlin: Difu Deutsches Institut für Urbanistik.
- Landes, B. (2011). Kostenmanagement in den Hilfen zur Erziehung. *Nachrichtendienst des deutschen Vereins für öffentliche und private Fürsorge*, (Juni), 277 - 281.

- Meinhold, M. (1996). *Qualitätssicherung und Qualitätsmanagement in der sozialen Arbeit*: Einführung und Arbeitshilfen. Freiburg im Breisgau: Lambertus.
- Moisl, D. (2011). Die Handlungsrelevanzmatrix als Baustein für die steuerungs- und entwicklungsorientierte Evaluation sozialer Dienstleistungen. *Zeitschrift für Evaluation*, Jg. 10 Heft 1, 69-97.
- Oelerich, G. (Hrsg.). (2005). *Soziale Dienstleistungen aus Nutzersicht: Zum Gebrauchswert sozialer Arbeit*. München and Basel: E. Reinhardt.
- Otto, H.-U. (2010). *Capabilities - Handlungsbefähigung und Verwirklichungschancen in der Erziehungswissenschaft*. Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH, Wiesbaden,.
- Otto, H.-U., & Ziegler, H. (o. J.). Der Capabilities-Ansatz als neue Orientierung in der Erziehungswissenschaft, *Otto, H.-U., & H. Ziegler. 2008. "Der Capabilities-Ansatz als neue Orientierung in der Erziehungswissenschaft", in: H.-U. Otto, & H. Ziegler (Eds.), Capabilities - Handlungsbefähigung und Verwirklichungschancen in der Erziehungswissenschaft. Wiesbaden: VS Verlag, 9 - 16.*
- Pruckner, M. (2005). *Die Komplexitätsfalle: wie sich die Komplexität auf den Menschen auswirkt*: vom Informationsmangel bis zum Zusammenbruch. Norderstedt: Books on Demand.
- Schmidt, C. M. (2010). Grundlagen evidenzbasierter Politikgestaltung bei sozialen Dienstleistungen. In Otto, H.-U. u.a. (Hrsg): *What Works - Welches Wissen braucht die Soziale Arbeit* (S. 165 - 183). Opladen & Farmington Hills: Verlag Barbara Budrich.
- Tornow, H. (2012). *WIMES Falldokumentation Manual*. Wülfrath: Eigenverlag des e/l/s-Institutes.
- Tornow, H., & Ziegler, H. (2012). Ursachen und Begleitumstände von Abbrüchen stationärer Erziehungshilfen (ABiE). *EREV Schriftenreihe*, 3(53. Jahrgang), 11 - 118.
- Walter Geiger; Willi Kotte. (2008). *Handbuch Qualität*: Grundlagen und Elemente des Qualitätsmanagements: Systeme - Perspektiven (5., vollst. überarb. und erw. Aufl.). Vieweg.

12 Anlage Glossar

Begriff	Erläuterung
Abbruch	vorzeitige Beendigung einer Hilfe im Dissens bei weiter bestehendem Bedarf; mindestens ein Partner steigt entgegen dem Wunsch/Rat der anderen aus einer Intervention aus; einseitiges Aufkündigen oder Nicht-Einhalten einer verabredeten Planung.
Benchmarking	Kennzahlen mit denen anderer Stichproben vergleichen; der „Benchmark“ ist die Vergleichsgröße.
Controlling	beobachtende und bewertende Aktivitäten in einem System, um Anhaltspunkte für steuernde Handlungen zu bekommen; Controlling ist nicht mit Kontrolle zu verwechseln, die ein Instrument der Machtausübung ist; Controlling ist eine organisationsinterne Dienstleistung für Entscheidungsträger und Verantwortliche, damit diese sich ein Bild von dem momentanen Zustand der Organisation und zu erwartenden Entwicklungen (==> Frühindikator) machen können.
Datenschutz	Rechtliche und technische Vorkehrungen, dass Sozialdaten nur für den Zweck verwendet werden, für den sie erhoben wurden. Datenschutz achtet auf Datensparsamkeit, informationelle Selbstbestimmung, Zugang zu Daten und die Zweckbindung der Verwendung; Daten müssen sicher vor Missbrauch und unerlaubtem Zugriff sein. Bürger müssen eine genaue Kenntnis darüber und einen Einfluss darauf haben, was mit ihren Daten passiert.
Datenschutz-Zertifikat	Bestätigung durch eine Unabhängige Datenschutzbörde, dass ein Verfahren rechtlich und technisch den Erfordernissen des Datenschutzes genügt; das WIMES-Web-Portal hat ein Datenschutzsiegel des ULD Kiel, der bundesweit die Eignung von datenverarbeitenden Verfahren für den öffentlichen Dienst prüft und zertifiziert.
Datensicherheit	Vorkehrungen gegen Datenverlust und für die Zugänglichkeit der Daten für die Zwecke, für die sie erhoben wurden; hängt eng mit dem ==> Datenschutz zusammen.
Design	hier im Zusammenhang mit „Untersuchungsdesign“, „Evaluationsdesign“: Welche Gruppen zu welchen Zeitpunkten verglichen werden.
Dialog	Verständigungsprozess zwischen Partnern, bei dem es nicht darum geht, dass ein Partner sich durchsetzt, sondern in dem in einer angstfreien, wertschätzenden und herrschaftsfreien Atmosphäre Verständnis füreinander wächst und neue, gute Lösungen für ein Problem gefunden werden.
Effektstärke (nach COHEN)	Grad eines erwünschten Effektes; der mittlere Unterschied wird in Bezug zu der Streuung der Anfangsmessung gesetzt ; wenn der Unterschied mehr als 50 % der Ausgangsstreuung beträgt, geht man von einer mittleren Effektivität aus; bei einem Unterschied von mehr als 80 % der Ausgangsstreuung geht man von einer hohen Wirksamkeit aus. Der Mittelwertunterschied muss statistisch signifikant sein.
EFQM	Qualitätsmanagementsystem der European Foundation For Quality Management; auch Bewertungssystem für die Exzellenz von Organisationen.
evidenzbasiert	auf nachgewiesenen wissenschaftlichen Erkenntnissen beruhend.
Frühindikator	==> Kennzahl, die zu einem frühen Zeitpunkt in einem Prozess auf Abweichungen und Risiken aufmerksam macht; Frühindikatoren dienen zur Steuerung von Prozessen, um möglichst genau zu erwünschten Ergebnissen zu kommen.

Begriff	Erläuterung
Grundgesamtheit	Gegenstand der Untersuchung oder Evaluation; bei der Evaluation besteht die Grundgesamtheit aus allen gegenwärtigen und zukünftigen Fällen des Systems, das evaluiert werden soll.
Handlungsziel	detailliertere Beschreibung eines Hilfeziels: "woran genau kann man erkennen, dass ein Ziel erreicht ist?" "Was genau macht bei gelingender Hilfe wer am Ende anders?"; ==> Richtungsziele werden in Handlungsziele spezifiziert und mit den Klienten vereinbart.
Indikator	siehe =Kennzahl
Kennzahl	quantitative Größe, die einen wesentlichen Aspekt in einem komplexen System abbildet; synonym: Indikator
Kooperation	Zusammenarbeit; Arbeitsteilung oder Arbeiten in einer Prozesskette: Der Output des einen ist der Input des anderen; wichtig sind ein vertrauensvoller Informationsaustausch und fehlerfreie Schnittstellen im Erstellungsprozess.
Korrelation	Zusammenhang zwischen zwei Messreihen (A und B); r kann Werte zwischen -1 und +1 annehmen; ein negatives Vorzeichen bedeutet: je höher A desto kleiner B; es gibt unterschiedliche Berechnungsarten, je nach der Art der Verteilungen.
Längsschnittuntersuchung	Vergleich mehrerer Messpunkte über die Zeit; z. B. am Anfang einer Intervention und am Ende bzw. zu einem Zeitpunkt nach dem Ende; ==> Querschnittuntersuchung
Makroebene	Unterstruktur in einem (sozialen) System, das die Einfluss auf das Gesamtsystem nimmt; z.B. in Organisationen die oberste Leitung
Median	deskriptive Kennzahl einer Verteilung von Messwerten; zentrales Moment; wenn man alle Messwerte aufsteigend sortiert, ist der mittlere Wert der Median; auch Perzentil 50: 50% der gemessenen Werte liegen unter dem Median; der Median ist von der Form der Verteilung unabhängig.
Mesoebene	Unterstruktur in einem (sozialen) System, die innerhalb des Gesamtsystems bestimmte Teilaufgaben übernimmt; z. B. regionale oder funktionale Differenzierungen in Organisationen
Mikroebene	kleinsten Elementen in einem (sozialen) Gesamtsystem; z.B. einzelne Mitglieder einer Organisation mit ihren Vernetzungen und Handlungen
Mittelwert (Mean)	deskriptive Kennzahl einer Verteilung von Messwerten; das arithmetische Mittel errechnet man, indem man alle Messwerte addiert und durch die Anzahl der Messwerte teilt. Mittelwerte setzen symmetrische Verteilungsformen voraus.
Nachhaltigkeit	Wirksamkeit über einen längeren Zeitraum; Vergleich zweier Messpunkte am Anfang einer Intervention und zu einem späteren Zeitpunkt nach dem Ende einer Intervention (==> Längsschnittuntersuchung)
online	in Echtzeit über Telekommunikationstechniken zum Informationsaustausch verbunden; die Webtechnologie für WIMES gestattet zum Beispiel dem Jugendamt und dem Leistungserbringer gleichzeitig einen Online-Datenzugang. Auch der Support kann den derzeitigen Datenstand in Echtzeit sehen, ohne geschützte Daten entschlüsseln zu können.
Pareto-Prinzip	80/20 Regel; entdeckt von dem Sozial- und Wirtschaftswissenschaftler A. Pareto im 19. Jh.; 80 % der Ereignisse verursachen 20 % der Kosten; die restlichen 20 % der Ereignisse verursachen 80 % der Kosten; mit anderen Worten: 80 % der Erfolge sind leicht zu erzielen; für die restlichen 20 % zur Perfektion braucht man 16 mal so viel Aufwand.

Begriff	Erläuterung
prospektive Evaluation	Gegenteil von ==> retrospektiver Evaluation; die Daten werden von Anfang an erhoben und für Vorhersagen genutzt (ex ante)
Qualitätsentwicklung	Aktivitäten zur Weiterentwicklung / Verbesserung eines Systems / einer Organisation; der Auftrag an die Leistungsträger, Jugendhilfe in ihrer Qualität zu entwickeln, ist gesetzlich verankert; allgemein bedeutet QE die kontinuierliche Verbesserung und das organisationale Lernen im Zusammenhang eines Qualitätsmanagementsystems
Qualitätssicherung	Aktivitäten zur Sicherstellung zugesagter Qualitätsmerkmale; dazu gehören Gewährleistungen der Strukturqualität, die stabile und zielführende Durchführung von Prozessen, die Erkennung und Beseitigung von Mängeln und Fehlern sowie die Messung der Ergebnisqualität. In der Jugendhilfe wird Qualitätssicherung skeptisch gesehen, weil sie in dem Verdacht steht, technokratisch zu sein und einer Standardisierung Vorschub zu leisten; aber auch hier gibt es im Gesetz und in Vereinbarungen die Erwartung, dass Mindeststandards eingehalten werden und Fehler frühzeitig erkannt und vermieden werden.
Querschnittuntersuchung	Vergleich zweier oder mehrerer Gruppen zu einem Zeitpunkt oder im gleichen Zeitraum; syn.: quasi-experimentelles => Design; => Längsschnittuntersuchung
retrospektive Evaluation	Die Daten für eine Evaluation werden post factum (ex post) nach dem Ergebnis zusammengetragen und bewertet.
Richtungsziel	Beschreibung des Bedarfs, bei dem eine Intervention etwas bewirken soll, auf mittlerer Abstraktionsebene; WIMES-Dimensionen, die in einer gezielten Hilfeplanung vorkommen, sind die Richtungsziele einer Hilfe.
sd	standard deviation; Standardabweichung; => Streuung
Signifikanz	Überzufälligkeit; zufällige Fallauswahl und Messfehler führen dazu, dass in Stichproben gemessene statistische Kennwerte, Zusammenhänge und Unterschiede nicht die Wirklichkeit widerspiegeln. Mit der Größe der Stichprobe und der Genauigkeit des Mess- und Berechnungsverfahrens steigt die Wahrscheinlichkeit, dass wahre Aussagen über die Grundgesamtheit gemacht werden und die Aussagen prognostischen Wert bekommen. Die statistische Signifikanz wird als Wahrscheinlichkeit angegeben, mit der eine Aussage aufgrund einer Stichprobe nicht den wahren Sachverhalt wiedergibt. $P>0,01$ heißt zum Beispiel, dass eine Aussage mit einer Wahrscheinlichkeit von unter 1 % falsch ist, weil sie auf Zufälle und Messfehler zurückzuführen ist; in den Sozialwissenschaften gibt man sich in der Regel mit $p>0,05$ zufrieden.
Sozialraumorientierung	Orientierungsrahmen sozialer Arbeit; Problemverständnis und Lösungsmöglichkeiten stehen immer im Sozialraumbezug; im erweiterten Sinne umfasst Sozialraumorientierung auch die Ausrichtung der Interventionen am Klientenwillen; auch die Steuerung der Ressourcen und der Einbezug der Träger kann über den Sozialraum erfolgen (Sozialraumbudgets, dezentrale Ressourcenverantwortung, kontraktierte Träger(verbünde)).
Stichprobe	zufällige Auswahl von Fällen aus einer Grundgesamtheit, bei denen ein Kennwert ermittelt wird oder ein Zusammenhang oder Unterschied untersucht werden soll; je größer die Stichprobe, desto höher die Verlässlichkeit (==> Signifikanz); Wenn Stichproben nicht zufällig ausgewählt werden, kann kein Rückchluss auf die Grundgesamtheit gemacht werden.

Begriff	Erläuterung
Streuung	Maß für die Unterschiedlichkeit der gefundenen Messwerte; die Einzelwerte "streuen" um einen Mittelwert; die Streuung ist gleich der Wurzel aus der ==> Varianz.
Systemdynamik	Veränderungen von Zuständen in einem System über die Zeit in ihrer Abhängigkeit von Veränderungen in der Umwelt oder innerhalb des Systems oder in Abhängigkeit von gezielten Reizen. Systemdynamik setzt ein Modell von wechselseitigen Beziehungen und von Regelkreisen voraus.
TQM	"total quality management"; am besten übersetzt mit "umfassendes Qualitätsmanagementsystem"; alle Aspekte einer Organisation werden in die Betrachtung mit einbezogen, verschiedene Perspektiven auf die Ergebnisse (kundenbezogen, mitarbeiterbezogen, gesellschaftsbezogen, geschäftsbezogen); bekanntestes System ==> EFQM
Varianz	Maß für die Unterschiedlichkeit der gefundenen Messwerte; alle Abstände zum Mittelwert werden quadriert, addiert und durch die Anzahl der Werte minus 1 geteilt. "Varianz" wird auch umgangssprachlich für die Unterschiedlichkeit von vorgefundene Merkmalsausprägungen gebraucht.
web-basiert	technische datenverarbeitende Systeme, die Teile der Technologie des Internets nutzen. Das WIMES-Web-Portal nutzt zum Beispiel die Browsetechnologie und den Transport der Daten über das Internet. Teile der Verschlüsselung sind ebenfalls der Internettechnologie entnommen (SSL-Verschlüsselung); andere Komponenten der Verschlüsselung und der Programmierung sind Eigenentwicklungen des els-Institutes.
WIMES	"Wirkung messen"; Methode zur Evaluation der ==> Wirksamkeit von Hilfen zur Erziehung
WIMES-Dimension	Zusammenfassung verwandter Bedarfe und Ziele; Lebensbereiche; Aufmerksamkeitsfelder Sozialer Arbeit; WIMES-Dimensionen fassen einzelne Aspekte eines Falles zusammen und liefern "Überschriften" für bestimmte Bedarfs- und Ziel-Cluster.
Wirksamkeit	Ausmaß, in dem eine bestimmte Interventionsart oder ein System in der Lage ist, Wirkungen hervorzubringen; syn.: Effektivität
Wirkung	hier: gewünschte Veränderung im Sinne einer Bedarfsminderung oder Zielerreichung, die mit ausreichender Plausibilität auf eine Intervention zurückzuführen ist.; syn.: Effekt
Wirkungsevaluation	wissenschaftliches Verfahren, die Wirksamkeit einer Interventionsart oder eines Systems zu bewerten; Wirkungsevaluation dient unmittelbar der Praxis und trifft immer nur Aussagen über den jeweiligen Untersuchungsgegenstand; Wirkungsevaluation will Praxis bewerten und Hinweise zur Qualitätsentwicklung geben.
Wirkungsforschung	Im Gegensatz zur Wirkungsevaluation will Wirkungsforschung allgemein gültige Erkenntnisse über ein Wissensgebiet erlangen und zur Modellbildung beitragen; im Vergleich zur => Wirkungsevaluation geht es um Theorie, Methode und allgemein gültige Aussagen.
Wirkungsfaktor	etwas, von dem die Wirkung abhängt; eine Eigenschaft des Systems oder eine Vorgehensweise, die geeignet ist, Wirkungen hervorzubringen.
Wirkungsindikator	Kennzahl, an der man Wirkung oder Wirksamkeit erkennen kann.



Bestimmung steuerungsrelevanter Wirkungsindikatoren HzE in Berlin (5 Bezirke)

Abschlussbericht für das Gesamtprojekt Wirkungsevaluation (2010 bis 2013)

Dr. Harald Tornow
e/l/s-Institut für Qualitätsentwicklung sozialer Dienstleistungen
Bernsstraße 7-9
42553 Velbert-Neviges
harald.tornow@els-institut.de
www.els-institut.de