

10. METHODEN

10.1. Übersicht über die Erhebungsebenen und Erhebungsarten

Zur möglichst umfassenden Programmevaluation eines Peer-Education-Programms zur Sexualaufklärung sind verschiedene Zielgruppen zu unterscheiden. Es sollen Programmwirkungen nicht nur auf der Ebene der Multiplikatoren (Peer-Educators), sondern auch bei den Adressaten (Zielgruppe) festgestellt werden. Dieses Interesse ist dadurch begründet, dass bisher so gut wie keine Ergebnisse aus experimentell oder quasiexperimentellen Untersuchungsansätzen zu Multiplikatoren- oder Adressatenwirkungen von Peer-Education-Programmen vorliegen. So ist es ein Anliegen, diese Lücke zumindest im Bereich der Adressatenwirkungen zu schließen. Es wird der Versuch unternommen - im Sinne umfassender Evaluation - Informationen zur Programmdurchführung und Programmwirkung zu erheben. Zur Validierung der Ergebnisse und zur zusätzlichen Erlangung von Prozessinformation wurden die Trainer mittels halbstandardisierter Interviews befragt. Die folgende Abbildung (Abbildung 10-1) stellt die drei Ebenen der Evaluation im Überblick dar: Die Ebene der Trainer, die Ebene der Multiplikatoren, sowie die Ebene der Adressaten.

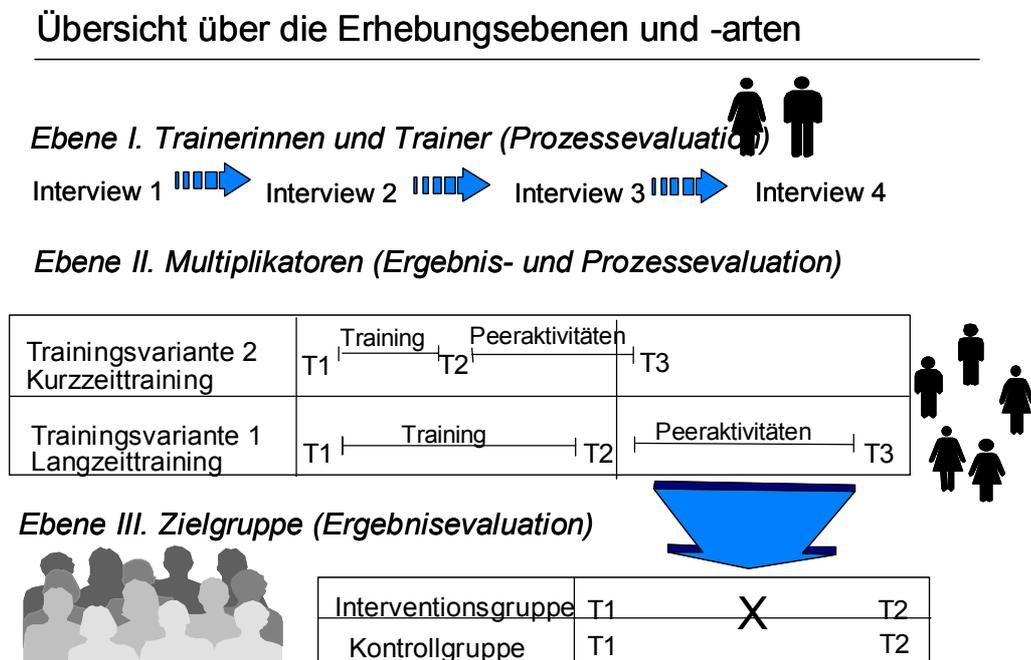


Abbildung 10-1. Überblick über Erhebungsebenen und -arten

Die Abbildung verdeutlicht auch die zeitliche Koordination der durchgeführten Untersuchungen. Als erstes wurden die Trainer interviewt, um mit ihrer Hilfe Erfolgskriterien entwickeln zu können. Als zweites folgte mit Beginn des Trainings die

Erstbefragung der Multiplikatoren und schließlich nach Abschluss des Trainings und vor Beginn der Durchführungsphase wurden Ausgangswerte der Adressaten erhoben. Die Informationsgewinnung war bezüglich Inhalt und Methode an die jeweilige Zielgruppe angepasst.

10.2. Die Erhebungsebene der Trainerinnen und Trainer

Um die subjektiven Zielsetzungen der Trainer in Bezug auf das Modellprojekt, ihre wahrgenommenen Erfolge und Misserfolge und generell auftretenden Implementationsprobleme verstehen und dokumentieren zu können, wurden problemzentrierte halbstrukturierte Interviews geführt. Die Mitglieder des pädagogischen Teams wurden in dem dreieinhalbjährigen Modellprojektverlauf insgesamt vier mal interviewt. Die qualitative Methode wurde vor geschlossenen Fragebogenmethoden bevorzugt, weil sie in besonderer Weise geeignet ist, die subjektiven Zielsetzungen der Trainer zu verdeutlichen.

Da im Rahmen der Programmentwicklung und -implementation Probleme auftreten können, die nicht vorhersehbar sind, sind hier offene Erhebungsmethoden zu bevorzugen und trafen auch bei den Trainern auf hohe Akzeptanz. Eine Auswertung dieser Interviews wurde während des Modellprojektverlaufs sukzessiv vorgenommen und als Prozessinformation zur Verbesserung und Unterstützung des Programms genutzt. Für die vorliegende Arbeit werden nur die vierten und letzten Trainerinterviews verwertet, die ein halbes Jahr vor Modellprojektende geführt wurden. Beschrieben wird die retrospektive Einschätzung des Trainings, die favorisierten Arbeitsweisen und Zielsetzungen der Trainer und vergleichende Bewertung der beiden Trainingsvarianten „Kurzzeittraining“ und „Langzeittraining“. Diese Informationen können zu einem besseren Verständnis der Entstehung von Programmwirkungen beitragen.

10.3. Die Erhebungsebene der Multiplikatoren (Peer-Educators)

10.3.1. DAS DESIGN

In dem Berliner Peer-Education-Programm wurden zwei verschiedene Trainingsvarianten realisiert – ein Langzeit- und ein Kurzzeittraining. Beide Trainingsvarianten unterschieden sich unter anderem durch die Dauer der zur Verfügung stehenden Trainingszeit. Während das Langzeittraining sich ungefähr mit 60 Trainingsstunden über ein dreiviertel Jahr bis ein Jahr erstreckte, wurden für das Kurzzeittraining nach der ursprünglichen Konzeption ca. 16 bis 20 Stunden anberaumt. Die geplante Trainingszeit sollte entweder auf zwei achtstündige oder auf vier vierstündige Termine verteilt werden. Das Kurzzeittraining war

als ein Trainingskonzept mit einer stärker aktionsorientierten Zielsetzung gedacht, die hauptsächlich in Jugend- und Freizeitheimen zur Anwendung kommen sollte.

Das Langzeittraining hingegen sah eine Umsetzung auf der Ebene von Schulen vor. Geplant war hier, Informationsveranstaltungen zu Liebe, Sexualität und Schwangerschaftsverhütung durch Peer-Educators innerhalb von Schulstunden durchzuführen. Die auszubildenden Jugendlichen wurden in Schulen rekrutiert, in denen sie später selbst aktiv werden sollten. Mit Einschränkung lassen sich hier die beiden Programmtypen peer-development (Kurzzeittraining) und peer-delivery (Langzeittraining) (vgl. Shiner, 1999) wiedererkennen.

	T 1 Vor Beginn des Peer- Education-Trainings	T 2 Nach Beendigung des Peer-Education-Trainings	T 3 Nach Durchführung der Peer-Educator Veranstaltungen
Trainingsgruppe 1 Berliner Langzeittraining	N = 68	Ca. ein dreiviertel Jahr nach T1 N = 28	N = 21
Trainingsgruppe 2 Berliner Kurzzeittraining	N = 57	Zwei bis vier Monate nach T1 N = 23	N = 16

Anmerkung: Zu T2 und T3 sind Peer-Educators nicht berücksichtigt, die nicht zu einem oder beiden vorhergehenden Messzeitpunkten einen Fragebogen ausfüllten.

Abbildung 10-2. Versuchsplan zur Peer-Educator-Erhebung; Vergleich des Berliner Langzeittrainings mit dem Kurzzeittraining

Abbildung 10-2 zeigt den Versuchsplan zur Peer-Educator Erhebung mit den beiden Trainingsvarianten Langzeittraining und Kurzzeittraining, die als unabhängiger Gruppierungsfaktor dienen. Mittels eines standardisierten Fragebogens wurden als abhängige Variable unter anderem Information zu den Bereichen sexuelle Erfahrung und Verhütungserfahrungen, themenbezogenes Wissen, wahrgenommene Kommunikationskompetenzen und Indikatoren des Wohlbefindens erhoben. Dieser Fragebogen wurde unmittelbar vor Beginn (T1) und nach Abschluss (T2) des Trainings, sowie nach einer Phase der Durchführung von Peeraktivitäten (T3) eingesetzt.

Das Ziel dieser Evaluation ist der Vergleich verschiedener Trainingsvarianten hinsichtlich ihrer Wirkung, somit handelt es sich um eine vergleichende Evaluation. Die Alternative zur vergleichenden Evaluation stellt die isolierte Evaluation dar, die eine Kontrollgruppe ohne Treatment erfordert. Das Ziel der isolierten Evaluation ist es, durch den Vergleich einer nichtbehandelten Kontrollgruppe mit einer Treatmentgruppe zu eruieren, ob ein Programm überhaupt eine Wirkung hervorruft. Hier jedoch geht es um den Vergleich der Trainingsvarianten: Langzeittraining versus Kurzzeittraining. Bei der vergleichenden Evaluation werden in Konkurrenz stehende Programme miteinander verglichen, wobei üblicherweise zumindest ein Programm seine (isolierte) Wirksamkeit bereits unter Beweis gestellt hat (vgl. Hager, 1998).

Allerdings bedingt die unterschiedliche Trainingsdauer systematische Fehlerquellen, die zu einer Bedrohung der Validität der Ergebnisse führten. Wie in Abbildung 10-2

dargestellt, waren die Zeiträume zwischen den einzelnen Messungen so unterschiedlich, dass spezifische Auswirkungen der Trainingsvarianten durch differentielle Entwicklungseffekte überdeckt sein konnten. Rivalisierende Erklärungsansätze zur Entstehung von Programmwirkungen wie Reifungseinflüsse spielen gerade im Jugendalter eine wichtige Rolle und können treatmentunabhängig zu Veränderungen in abhängigen Variablen führen. Zum zweiten war die Vergleichbarkeit durch die unterschiedliche Geschlechtsverteilung der beiden Trainingsvarianten in Frage gestellt. Während bei dem Langzeittraining die Mädchen dominierten (62.7 Prozent Mädchen) nahmen bei dem Kurzzeittraining mehr Jungen teil (57.1 Prozent Jungen). Der Chi²-Test wird signifikant mit $\text{Chi}^2(1) = 4.824$, $p = .028$. Das Durchschnittsalter in beiden Gruppen war hingegen vergleichbar mit 14.7 Jahren bei Langzeittraining und 14.9 Jahren beim Kurzzeittraining.

Entgegen der ursprünglichen Absicht ein und denselben Fragebogen bei beiden Trainingsvarianten einzusetzen, musste der Fragebogen für das Kurzzeittraining stark gekürzt werden. Dies wurde notwendig, da sich in der praktischen Durchführung herausstellte, dass das Ausfüllen des Fragebogens vor Beginn des Trainings zu viel Zeit erforderte, die dann beim eigentlichen Training fehlte. Dies schränkte die Möglichkeit des Vergleichs zwischen Intensiv- und Kurzzeittraining noch weiter ein.

Daher wurde die Gelegenheit zu einem weiteren Vergleich des Langzeittrainings mit einem Kurzzeittraining genutzt (vgl. Abbildung 10-3). Zeitgleich zu dem Berliner Peer-Education-Programm wurde in Graz (Österreich) ein weiteres Peer-Education-Programm zur HIV-Prävention durchgeführt. In Graz wurden in Absprache mit dem Berliner Modellprogramm die gleichen Messinstrumente wie in dem Berliner Langzeittraining eingesetzt, daher wird sich der Vergleich auf das Langzeittraining beschränken. Das Grazer Training ist auf ein bis zwei Intensivwochenenden beschränkt und wird sehr strukturiert durchgeführt. Im Unterschied zu dem Berliner Kurzzeittraining wurde das Grazer Programm nicht an Freizeiteinrichtungen, sondern ebenso wie das Berliner Langzeittraining an Schulen durchgeführt.

	T 1	Trainings- und Durchführungsphase	T 3
Trainingsgruppe 1 Berliner Langzeittraining	N = 69	1 ½ Jahre	N = 21
Trainingsgruppe 3 Grazer Kurzzeittraining	N = 31	½ Jahr	N = 28
Kontrollgruppe 3 Zum Grazer Kurzzeittraining	N = 31	4 Monate	N = 31

Abbildung 10-3. Versuchsplan zur Peer-Educator Erhebung; Vergleich mit dem Grazer Kurzzeittraining

Das Grazer Programm verfügt über den Vorteil einer Kontrollgruppe. Diese kann zwar nicht als Kontrollgruppe für das Berliner Langzeittraining dienen, bietet aber immerhin die Möglichkeit, dass die isolierte Wirksamkeit des Grazer Kurzzeitprogramms unter Beweis

gestellt werden kann. So kann beispielsweise der Einfluss von Reifungsprozessen über einen Zeitraum von vier Monaten abgeschätzt werden.

10.3.2. DIE DURCHFÜHRUNG DER BEFRAGUNG

Die Multiplikatoren wurden zu Beginn des Modellprogramms zu Gruppen zu jeweils ca. 8 bis 12 Jugendlichen zusammengefasst. Die Befragung der Jugendlichen erfolgte sowohl beim Langzeittraining, als auch beim Kurzzeittraining für jede Peer-Education Gruppe getrennt. Die Gruppen begannen zeitlich versetzt mit ihrer Ausbildung, so dass sich Trainings- und Durchführungsaktivitäten und damit auch die Befragungen über einen Zeitraum mit Beginn der ersten Schulung zwischen Mai 1995 und November/Dezember 1997 mit Durchführung der letzten Follow-up Befragungen erstreckten. Die Befragungen wurden von studentischen und wissenschaftlichen Projektmitarbeitern in den Räumen der Senatsverwaltung für Gesundheit und Soziales¹, bzw. beim Kurzzeittraining in den Räumen der Freizeiteinrichtung durchgeführt. Die Räumlichkeiten waren identisch mit dem Ort der Trainingsdurchführung. Um beiden Geschlechtern gleichermaßen die Möglichkeit für Nachfragen zu bieten, wurde zu Beginn darauf geachtet, dass die Befragungen durch Projektmitarbeiter beiderlei Geschlechts durchgeführt wurden. Bei späteren Befragungen war dies aus Gründen mangelnder personeller Kapazität nicht mehr möglich. Die Befragungen waren recht umfangreich und dauerten zwischen einer und eineinhalb Stunden.

Die erste Befragung (T1) erfolgte jeweils unmittelbar im Anschluss an eine drei bis vierstündige Einführungsveranstaltung („Schnupperveranstaltung“) der Trainer. Die zweite Befragung (T2) wurde zum Trainingsabschluss durchgeführt. Als Trainingsabschluss wurde der Zeitpunkt definiert, zu dem die Jugendlichen sich für die Durchführung einer Informationsveranstaltung oder anderen Peeraktivität entschieden und diese vorbereitet hatten. Beim Langzeittraining beinhaltet die Trainingsdauer im Durchschnitt etwa ein dreiviertel Jahr. Nach diesem Zeitpunkt konnten noch Treffen mit den Trainern zur Vorbereitung oder Nachbereitung der Aktion stattfinden. In keinem Fall jedoch hatte schon vor der Befragung eine Peeraktion stattgefunden. Die Follow-up-Befragung (T3) wurde ein halbes bis dreiviertel Jahr nach der zweiten Befragung durchgeführt. Während dieses Zeitraums hatten die Peer-Educators die Gelegenheit, Aktionen innerhalb oder außerhalb von Schulklassen durchzuführen.

Kapitel 13 thematisiert Veränderungen und Erfahrungen der Multiplikatoren mit Langzeittraining, die sowohl im Verlaufe des Trainings, als auch der sich anschließenden Phase eigener „Peeraktivitäten“ beschreibbar sind.

In Kapitel 14 geht es um die spezifischen Erfahrungen und Aktivitäten der Jugendlichen mit Kurzzeittraining, die - soweit möglich - zu Jugendlichen mit Langzeittraining in Beziehung gesetzt werden.

¹ Die Senatsverwaltung wurde in der Zwischenzeit umstrukturiert

10.3.3. MESSINSTRUMENTE

10.3.3.1. *Der Fragebogen für Peer-Educators (Langzeittraining)*

Der Fragebogen zur Erfassung der Programmwirkungen bei Multiplikatoren wurde von Kleiber und Pforr entwickelt und vor Beginn der Evaluationsstudie auf Verständlichkeit und Verteilungseigenschaften der Items überprüft. Eine ausführliche Dokumentation aller Items mit Quellenangabe ist im Anhang abgebildet. Dargestellt sind Mittelwerte der Items und der Skalen, Standardabweichung, Minimum, Maximum und Schiefe. Darüber hinaus findet sich Information zur Reliabilität zum ersten (T1), zweiten (T2) und letzten (T3) Messzeitpunkt, sowie zur Korrelation zwischen den Messzeitpunkten.

Die Motive zur Teilnahme an dem Modellprogramm könnten eine zentrale Bedingung des Programmerfolgs sein. Wenn Programmziele mit Bedürfnissen und Zielen der Betroffenen nicht kompatibel sind, können hohe Dropout-Quoten oder fehlende Programmwirkungen die Folge sein. Mögliche Beweggründe Jugendlicher zur Teilnahme an einem Peer-Education-Programm wurden anhand einer offenen Frage in einem Pretest erfragt. Die Antworten dienten als Basis für die Konstruktion einer Anzahl von Items, die bei der Erstbefragung der Multiplikatoren eingesetzt wurden. Ausgehend von diesem Itempool von insgesamt 26 Items wurden faktorenanalytisch drei Faktoren gewonnen. Die drei resultierenden Skalen waren „Altruismus“ (Cronbach's $\alpha = .84$), „Wunsch nach Anerkennung“ (Cronbach's $\alpha = .73$) und „Informationssuche“ (Cronbach's $\alpha = .71$).

Das themenspezifische Wissen wurde durch eine insgesamt 20 Items umfassende Skala in Anlehnung an Eichholz, Niehammer, Wendt und Lohaus (1994) erfasst. Für jede der Fragen waren vier Antwortmöglichkeiten vorgegeben. Nur eine der Antwortmöglichkeiten war richtig. Von der ursprünglich 24 Items umfassenden Skala von Eichholz et al. (1994) wurden 17 Items übernommen (davon wurde ein Item modifiziert). Die Skala wurde durch drei selbst entwickelte Items, die HIV und Kondomgebrauch thematisierten, ergänzt. Die Jugendlichen konnten daher einen maximalen Gesamtwert von 20 richtig gelösten Aufgaben erreichen.

Kondomakzeptanz wurde in Anlehnung an Oswald und Pforr (1991) erfragt. Die 13-Item Skala wurde faktorenanalytisch auf eine 5-Item Skala reduziert und als „Negative Kondomeinstellung“ bezeichnet. Ein niedriger Wert auf dieser Skala ist also mit einer positiven Einstellung gegenüber Kondomen gleichzusetzen. Aufgrund des vierstufigen Abfragemodus waren Mittelwerte zwischen 1 und 4 zu erreichen. Die Reliabilität zu T1 ist mit Cronbach's $\alpha = .77$ als zufriedenstellend zu kennzeichnen, sinkt über die Messzeitpunkte jedoch rapide auf Cronbach's $\alpha = .52$ zu T2 und Cronbach's $\alpha = .05$ zu T3. Daher wurde im weiteren Verlauf auf eine Auswertung der Kondomakzeptanz verzichtet, mit Ausnahme der Dropoutanalysen. Hier ging nur die Kondomakzeptanz zum ersten Messzeitpunkt in die Auswertung mit ein.

Die Kommunikationsschwierigkeit wurde durch Selbsteinschätzungen bezüglich des Umgangs mit kommunikativ schwierigen Situationen erfragt (Beispielitem: „Wie schwierig wäre es im Moment für Dich über heikle Themen, bzw. Sexualität vor einer Gruppe zu reden“). Die 12 – Item Skala erwies sich in einer Faktorenanalyse als eindimensional und verfügte über eine hohe interne Konsistenz (Cronbach's α zu T1 = .89), die auch zu T2 (Cronbach's α = .80) und T3 (Cronbach's α = .78) bestehen blieb. Durch den sechsstufigen Abfragemodus konnten Mittelwerte zwischen 1 und 6 erreicht werden. Aufgrund der erforderlichen Selbsteinschätzung schwieriger kommunikativer Situationen kommt diese dem Konstrukt „kommunikationsspezifische Selbstwirksamkeitserwartung“ nahe. In der späteren Ergebnisdarstellung wird die Kommunikationsschwierigkeit umgepolt und als Kommunikationskompetenz bezeichnet.

Wohlbefinden wurde durch den Berner Fragebogen zum Wohlbefinden Jugendlicher (Grob, Lüthi, Kaiser, Flammer, Mackinnon & Wearing, 1989) erfragt. Da dieser Fragebogen an einer Stichprobe Schweizer Jugendlicher entwickelt und darüber hinaus kreuzvalidiert wurde, sollte die Bildung der Skalen inhaltlich – theoretisch in Anlehnung an Grob et al. (1989) erfolgen. Der Fragebogen umfasst insgesamt 36 Items mit den sechs Faktoren „Positive Lebenseinstellung“ (sechsstufig), „Problembewusstheit“ (fünfstufig), „Körperliche Beschwerden“ (vierstufig), „Selbstwert“ (sechsstufig), „Depressive Stimmung“ (sechsstufig) und „Lebensfreude“ (vierstufig). Die internen Konsistenzen waren jedoch nicht in jedem Fall zufriedenstellend, so dass eine Erhöhung der Konsistenzen mittels Elimination einzelner Items versucht wurde. Eine Verbesserung gelang jedoch nur für die Skala „Selbstwert“, die auf drei Items verkürzt werden musste. Zu T1 finden sich somit die folgenden Reliabilitäten: „Positive Lebenseinstellung“ (Cronbach's α = .83), „Problembewusstheit“ (Cronbach's α = .65), „Körperliche Beschwerden“ (Cronbach's α = .74), „Selbstwert“ (Cronbach's α = .81), „Depressive Stimmung“ (Cronbach's α = .65) und „Lebensfreude“ (Cronbach's α = .73). Weitere Information zur Reliabilität und Stabilität der Subskalen findet sich im Anhang.

Zur Beurteilung der internen Validität des Berner Fragebogens zum Wohlbefindens werden Interkorrelationen der Subskalen des Fragebogens berechnet. Die Interkorrelationstabelle (vgl. Tabelle 10-1) verweist auf mittlere bis hohe gleichsinnige Zusammenhänge zwischen „Positive Lebenseinstellung“, „Selbstwert“ und „Lebensfreude“. Je ausgeprägter die Werte auf allen drei Skalen, desto geringer die depressive Verstimmung, wie die jeweils mittlere negative Korrelation zeigt. Auffällig ist der hohe Zusammenhang zwischen „Körperliche Beschwerden“ und „Problembewusstheit“. Offensichtlich berichten Jugendliche mit ausgeprägterem Problembewusstsein auch mehr körperliche Beschwerden. Dies kann als ein Hinweis gewertet werden, dass ein Teil der Ergebnisse bei körperlichen Beschwerden auch durch die Bereitschaft Jugendlicher bedingt ist, Beschwerden wahrzunehmen und zu berichten.

Tabelle 10-1. Interkorrelation der Subskalen des Berner Fragebogens zum Wohlbefinden Jugendlicher zum ersten Messzeitpunkt

	1	2	3	4	5	6
1 Positive Lebenseinstellung						
2 Problembewusstheit	-.22 *					
3 Körperliche Beschwerden	-.15ns.	.63***				
4 Selbstwert	.58***	-.28*	-.15ns.			
5 Depressive Verstimmung	-.49***	.22*	.02ns.	-.47***		
6 Lebensfreude	.46***	-.03ns.	.05ns.	.36***	-.45***	

Zur Bestimmung des „sexuellen Selbstbildes“ wurde eine Skala eigens für Forschungszwecke konstruiert. Diese Skala umfasste ursprünglich 13 teilweise recht persönliche Aussagen zur Sexualität, zum Beispiel „In bezug auf meine Sexualität fühle ich mich sicher“ oder „Ich glaube, dass ich sexuelle Probleme habe“. Eine Faktorenanalyse ergab nur einen inhaltlich validen und reliablen Faktor (Cronbach's α zu T1 = .80), der „Sexuelle Selbstsicherheit“ genannt wurde. Allerdings ging auch hier die Reliabilität zu T2 (Cronbach's α = .58) und T3 (Cronbach's α = .67) zurück, was bei der weiteren Interpretation der Ergebnisse zu berücksichtigen ist.

Zusätzlich zu den beschriebenen Skalen wurden auch offene Fragen gestellt, um den Jugendlichen die Möglichkeit zu geben, ihre persönliche Sichtweise zu Auswirkungen des Programms zu äußern. So wurden bei der letzten Follow-up Befragung drei offene Fragen mit folgendem Wortlaut gestellt: (1) „Was hat es (das Programm) Dir persönlich an Erfahrungen gebracht oder was hat sich dadurch für Dich geändert?“ (2) „Welche Erfahrung empfundest Du dabei als ganz besonders wichtig oder schön?“ (3) „Gab es auch etwas, das ganz besonders schwierig oder unangenehm für Dich war?“ Die Antworten dienen dazu, den Programmerfolg zu illustrieren und Hinweise zur Validität des quantitativen Befragungsteils zu bekommen. Außerdem sollten sie auch Erklärungen für Programmerfolge oder Umsetzungsprobleme aus Sicht der Jugendlichen geben.

10.3.3.2. Der Fragebogen für Peer-Educators (Kurzzeittraining)

Aufgrund des zu hohen zeitlichen Aufwandes beim Ausfüllen des Fragebogens in Relation zur zur Verfügung stehenden Trainingszeit musste der Fragebogen zum Kurzzeittraining stark reduziert werden. Einerseits wurden Skalen gekürzt, andererseits wurden Skalen zu T2 eliminiert. Dies schränkt die Möglichkeiten des Vergleichs zwischen Langzeit- und Kurzzeittraining (vgl. Abbildung 10-2) ein. Detaillierte Information zu Items und gebildeten Skalen zum Kurzzeittraining ist im Anhang abgebildet.

Das themenspezifische Wissen wurde wie beim Langzeittraining nach Eichholz et al. (1994) erfasst, allerdings wurde die insgesamt 20 Items umfassende Skala auf 12 Items reduziert. Die Erfassung der „Kommunikationsschwierigkeit“, bzw. „Kommunikationskompetenz“ war identisch mit dem Langzeittraining. Die interne Konsistenz über alle drei Messzeitpunkte zufriedenstellend von T1 (Cronbach's α = .79) über T2 (Cronbach's α = .77) und T3 (Cronbach's α = .90). Aus dem Berner Fragebogen zum Wohlbefinden wurden drei Subskalen verwendet „Positive Lebenseinstellung“,

„Körperliche Beschwerden“ und „Selbstwertgefühl“. Diese wurden identisch mit dem Langzeittraining erfragt, jedoch nur zu T1 und T3 eingesetzt. Die internen Konsistenzen lagen beispielsweise bei Selbstwertgefühl für T1 bei Cronbach's $\alpha = .60$ und zu T2 bei Cronbach's $\alpha = .69$ und waren damit insgesamt zu niedrig (vgl. Anhang). Die gesamten Skalen zur Kondomakzeptanz und zum sexuellen Selbstbild entfielen.

10.3.3.3. Der Fragebogen für Peer-Educators (Grazer Programm)

Dieser Fragebogen ist weitestgehend identisch mit dem Fragebogen, der in dem Berliner Langzeittrainingsprogramm eingesetzt wurde. Da das Training auf ein bis zwei Wochenenden beschränkt war, wurde der ausführliche Fragebogen nur zum ersten (T1) und letzten (T3) Messzeitpunkt eingesetzt. Unmittelbar nach dem Training wurde nur Information zur Bewertung des Trainings und des Leitungsstils der Trainer erhoben. Zur Erfassung des themenspezifischen Wissens wurden leichte Modifikationen vorgenommen, da ein wesentliches Ziel des Grazer Trainings in der Vermittlung HIV-relevanten Wissens bestand. Item 20, das sich auf HIV-Wissen bezog, wurde eliminiert, dafür weitere Fragen zum Wissen ergänzt, die jedoch in der vorliegenden Auswertung nicht berücksichtigt werden konnten. Um Auswirkungen beider Trainings (vgl. Abbildung 10-3) vergleichen zu können, wurde die Skalenbildung des Berliner Langzeitprogramms auf das Grazer Programm übertragen. Zum Vergleich des themenspezifischen Wissens konnten allerdings nur 19 Items berücksichtigt werden. Analog zu den Berliner Fragebogen ist detaillierte Information zu einzelnen Skalen im Anhang abgebildet. Folgende Tabelle gibt einen Überblick über die internen Konsistenzen der gebildeten Skalen zum ersten und dritten Messzeitpunkt.

Tabelle 10-2. Interne Konsistenzen (Cronbach's α) der Skalen des Grazer Kurzzeitprogramms

Skalen	Interne Konsistenz T1	Interne Konsistenz T3
Teilnahmemotive ^a		
„Altruismus“	$\alpha = .76$	---
„Anerkennung“	$\alpha = .62$	---
„Informationssuche“	$\alpha = .66$	---
Wohlbefinden		
„Positive Lebenseinstellung“	$\alpha = .72$	$\alpha = .74$
„Problembewusstheit“	$\alpha = .70$	$\alpha = .62$
„Körperliche Beschwerden“	$\alpha = .67$	$\alpha = .72$
„Selbstwert“	$\alpha = .77$	$\alpha = .79$
„Depressive Stimmung“	$\alpha = .80$	$\alpha = .74$
„Lebensfreude“	$\alpha = .58$	$\alpha = .74$
Spezifische personale Ressourcen		
„Kommunikationsschwierigkeit“	$\alpha = .84$	$\alpha = .88$
„Negative Kondomeinstellung“	$\alpha = .58$	$\alpha = .62$
„Sexuelle Selbstsicherheit“	$\alpha = .71$	$\alpha = .58$

Anmerkung: Teilnahmemotive wurden nur zum ersten Messzeitpunkt (T1) erfragt.

10.3.4. DIE BESCHREIBUNG DES TRAININGS

Das Training der Peer-Educators folgt keinem vollständig festgelegten Programm, da ansonsten aus Sicht der Trainer die Partizipationsmöglichkeiten der Peer-Educators stark minimiert worden wären. Das pädagogische Team hatte eine Reihe von Trainingskomponenten vorgesehen, die von den Jugendlichen gewählt werden konnten. Potentielle Schulungsinhalte waren (nach Schönbach, 1995):

- Wissensvermittlung zu den Themenfeldern: Verhütungsmittel, Geschlechtskrankheiten, AIDS, HIV, Testproblematik, Betroffenenproblematik, Infektionswege, Pubertät, Selbstbefriedigung, Homosexualität, Gefühle, Erotik usw.
- Werte, Normen, Rollenverhalten
- Fragen zu Treue, Trennung, Partnerschaft und Beziehungen

Methodisch wurden Kurzvorträge, Selbsterfahrungsmethoden, Kommunikationstechniken und z.B. Filme eingesetzt. Darüber hinaus bestand die Möglichkeit von Besuchen der Berliner AIDS-Hilfe und eines Selbsthilfevereins für Homosexuelle (Mann-O-Meter e.V.) Das Einbeziehen von solchen Beratungseinrichtungen dient auch dazu, diese kennenzulernen und Schwellenängste abzubauen. Ursprünglich war ein ca. halbjährliches Training angestrebt, in der Praxis zog sich die Ausbildung der Peer-Educators bis zu einem Jahr hin. Teilweise war die Verzögerung durch Terminprobleme bedingt, teilweise hatten die Peer-Educators subjektiv das Gefühl noch mehr Vorbereitung zu benötigen. Eine laufende Prozessevaluation der einzelnen Trainingssitzungen mit laufender Beschreibung von Trainingsinhalten durch Pädagogen und nachfolgender Bewertung durch Multiplikatoren wurde von den Trainern als nicht praktikabel angesehen. Jedoch wurde in

allen Kurzzeit- und Langzeittrainings unmittelbar nach Abschluss des Trainings Einschätzungen der Peer-Educators bezüglich des Trainings und des Leitungsstils der Trainer erfragt.

10.3.4.1. Fragen zur Bewertung des Leitungsstils und der Trainermethoden

Die Ergebnisse der Führungsstilforschung gaben der Entwicklung des gruppenpädagogischen Konzeptes in den USA und in Deutschland entscheidende Anstöße. Lewin war der erste, der gemeinsam mit seinen Mitarbeitern die Wirkung der Führungsstile von Gruppenleitern auf das soziale Verhalten von Menschen und die Gruppenatmosphäre erforschte. Er entwickelte die bekannte Typologie „Autoritärer Führungsstil“, „Demokratischer Führungsstil“ und „Laizzes - Faire Führungsstil“. Vor allem der „demokratische Führungsstil“ findet sich in Prinzipien der Gruppenpädagogik wieder (vgl. Sader, 1991). Es war nicht beabsichtigt, diese Typologien in unseren Messinstrumenten abzubilden, sondern lediglich einige Aspekte des Leitungsstils genauer zu beleuchten, die eine Bewertung des pädagogischen Konzeptes der Trainer erlauben.

Insgesamt 15 Items, die Bewertungen und Aussagen zum Leitungsstil der Trainer erfassen sollten, wurden mit Hilfe einer exploratorischen Faktorenanalyse auf vier zugrundeliegende Dimensionen reduziert (vgl. Tabelle 10-3). Die gefundene Faktorenstruktur ist aufgrund der geringen Stichprobenzahl nicht als stabil zu betrachten, sondern spezifisch auf unsere Stichprobe zugeschnitten und dient in erster Linie einer Erhöhung der Reliabilität und Strukturierung der Ergebnisse. Der erste Faktor klärt 25 Prozent Varianz auf und beinhaltet Aussagen wie „die Trainer ... können offen über Sexualität reden“, „... drücken sich verständlich aus“. Dieser Faktor wurde als „didaktisch-kompetenter Leitungsstil“ (Cronbach's $\alpha = .75$) bezeichnet. Der zweite Faktor war gekennzeichnet durch eine stärkere Leitungszentrierung, die mit der Bevorzugung einzelner Jugendlicher und einer Dominanz von Wissensvermittlung einherging. Dieser Leitungsstil wurde als „expertenorientierter Leitungsstil“ bezeichnet. Zur Bildung der Skala wurde das Item „... sind mir zu kumpelhaft“ ausgeschlossen, da es inhaltlich in einem gewissen Widerspruch zu der gewählten Skalenbezeichnung stand. Dadurch verbesserte sich auch die interne Konsistenz der Skala (Cronbach's $\alpha = .85$). Der dritte Faktor war inhaltlich durch eine Anpassung an Bedürfnisse und Probleme der Gruppe gekennzeichnet. Dieser Faktor wurde „adaptiver Leitungsstil“ (Cronbach's $\alpha = .70$) benannt. Der vierte und letzte Faktor konnte als Skala nicht berücksichtigt werden, da die Bildung einer Skala eine Mindestzahl von drei Items voraussetzt. Die drei Faktoren sind aufgrund der gewählten schiefwinkligen Rotationsmethode nicht unabhängig voneinander.

Tabelle 10-3. Faktorenanalyse der Leitereinschätzungen durch Peer-Educators (N = 59)

Die Gruppenleiter ...	Faktor 1	Faktor 2	Faktor 3	Faktor 4	Kommunalitäten
... können offen über Sexualität reden	.79				.73
... drücken sich verständlich aus	.77				.67
... wissen viel zum Thema Sexualität	.71				.52
... finde ich toll	.68				.70
... nehmen mich ernst	.68				.52
... bevorzugen manchmal einzelne Gruppenteilnehmer		.87			.77
... nehmen die reine Wissensvermittlung zu wichtig		.86			.83
... lassen in der Gruppe zu wenig Freiraum für unsere eigenen Ideen		.78			.62
... sind mir zu lehrerhaft		.65			.54
... sind mir zu kumpelhaft		.62			.51
... nehmen sich genügend Zeit			.79		.70
... können mit Konflikten in der Gruppe gut umgehen			.77		.60
... sorgen dafür, dass alle sich wohl fühlen			.66		.46
... lassen jeden zu Wort kommen				-.60	.68
... können sich gut in mich hineinversetzen				.51	.46
Eigenwert	4.29	2.64	1.54	1.01	
Varianzaufklärung	28,6 %	17,7 %	10,3 %	6,8 %	63,4 %

Anmerkung. FA nach Hauptkomponentenmethode mit schiefwinkliger Rotation;

Die Faktorenanalyse basiert auf einer Stichprobe von 59 Peer-Educators aus beiden Trainingsvarianten, die das Training beendet haben. Hier besteht eine Differenz zur Abbildung 10-2, aus der ersichtlich wird, dass nur 51 Peer-Educators zum zweiten Messzeitpunkt befragt wurden. Die Differenz von 8 Peer-Educators ergibt sich dadurch, dass diese offensichtlich verspätet in das Training eingestiegen sind, den ersten Fragebogen nicht ausgefüllt haben und daher bei der Gesamtauswertung der Programmwirkungen nicht berücksichtigt werden konnten. Um zu möglichst stabilen, vom Zufall weitgehend unbeeinflussten Faktorenstrukturen zu gelangen, sollte nach Bortz (1984) die untersuchte Stichprobe möglichst groß und repräsentativ sein. Daher wurden diese acht Peer-Educators speziell in dieser Analyse berücksichtigt. Nach Bortz (1984, S. 630) ist zumindest darauf zu achten, dass die Anzahl der Untersuchungseinheiten die Anzahl der zu faktorisierenden Merkmale übersteigt. Dies war bei der durchgeführten Faktorenanalyse der Fall. Die folgende Tabelle (vgl. Tabelle 10-4) zeigt Interkorrelationen der drei gebildeten Skalen, sowie deren Korrelation zum Gruppenklima (Cronbach's $\alpha = .70$). Das Gruppenklima wurde durch die drei Items erfasst: „Ich konnte meine Gedanken und Meinungen in die Gruppe einbringen.“, „Meine Gedanken und Meinungen wurden von den anderen respektiert“ und „Ich fühlte mich in der Gruppe gut aufgehoben.“

Der expertenorientierte Leitungsstil korreliert negativ mit dem adaptiven Leitungsstil ($r = -.35$), was inhaltlich plausibel erscheint, da beide Leitungsarten im Widerspruch zueinander stehen. Zwischen expertenorientiertem Leitungsstil und Wahrnehmung der Trainer als didaktisch und fachlich kompetent scheint allerdings kein Zusammenhang zu bestehen. Im Gegensatz dazu jedoch bilden sich Zusammenhänge zwischen der Einschätzung des Leitungsstils als didaktisch-kompetent und dem adaptiven Leitungsstil

ab. Je stärker der Gruppenleitungsstil an den individuellen Bedürfnissen der Gruppe orientiert war, desto didaktisch-kompetenter wurden die Trainer wahrgenommen.

Alle drei Leitungsstilarten korrelieren wie zu erwarten mit dem Gruppenklima, was für die Validität der gebildeten Skalen spricht. Der expertenorientierte Leitungsstil korreliert im Gegensatz zu den anderen Skalen negativ mit dem Gruppenklima.

Tabelle 10-4. Interkorrelationen der Leitungsstilarten (N = 59)

	Gruppenklima	Leitexp	Leitadap	Leitkomp
Gruppenklima				
Leitexp	-.35**			
Leitadap	.32*	-.35**		
Leitkomp	.31*	-.15ns.	.40 **	

Anmerkung. Leitexp = expertenorientierter Leitungsstil, Leitadap = adaptiver Leitungsstil, Leitkomp = kompetenter Leitungsstil

Daraus könnte man schließen, dass entsprechend den gruppenpädagogischen Arbeitsprinzipien, ein adaptiver Leitungsstil für das Entstehen von Vertrauen und Offenheit in der Gruppe günstig ist, ein expertenorientierter Leitungsstil sich hingegen eher hemmend auf die Entwicklung eines offenen Gruppenklimas auswirkt. Prinzipiell sind auch umgekehrte Zusammenhänge und Wechselwirkungen zwischen Leitungsstil und Gruppe nicht auszuschließen. Eine vertrauensvolle und offene Gruppenatmosphäre mag den Pädagogen erst die Möglichkeit geben, Kontrolle und Steuerung zurückzunehmen und sich auf die Wünsche und Bedürfnisse der Gruppenteilnehmer einzulassen.

Einschätzung des Trainings

Die Jugendlichen sollten in insgesamt 14 Aussagen themen- und gesprächszentrierte, erlebnis- und wahrnehmungsbezogene Teilaspekte des Trainings einschätzen. Eine Faktorenanalyse erbrachte allerdings keine inhaltlich-sinnvoll interpretierbare Faktoren mit eindimensionaler Ladung. Daher wurden die Items anhand inhaltlich-theoretischer Überlegungen auf drei Skalen reduziert. Skala 1 besteht aus 3 Items und beinhaltet das „Überschreiten persönlicher Grenzen“ (Cronbach's $\alpha = .79$). Skala 2 wird aus 4 Items gebildet und wird als „Thematisierung von Sexualität und Verhütung“ bezeichnet (Cronbach's $\alpha = .87$). Skala 3 schließlich besteht ebenfalls aus 4 Items und wird als „Sensibilisierung für Selbst- und Fremdwahrnehmung“ (Cronbach's $\alpha = .84$) etikettiert. Eine Interkorrelation der Skalen zeigt allerdings einen hohen Zusammenhang der beiden Subskalen „Sensibilisierung von Selbst- und Fremdwahrnehmung“ sowie „Thematisierung von Sexualität und Verhütung“ mit einer Korrelation von $r = .76$. Auf der Konstruktebene sind die beiden Skalen nicht voneinander zu trennen. Aufgrund der unterschiedlichen Inhalte jedoch wird die Trennung der beiden Skalen für die nachfolgende Auswertung beibehalten. Der Wortlaut der Items und weitere statistische Information zu Mittelwerten und Verteilung der Items bezogen auf Langzeittraining und Kurzzeittraining ist im Anhang dargestellt.

10.4. Die Erhebungsebene der Adressaten

10.4.1. DAS DESIGN

Während bei der Feststellung von Programmwirkungen bei Multiplikatoren eine vergleichende Evaluation mit verschiedenen Treatmentgruppen (Langzeittraining, Kurzzeittraining) durchgeführt wurde, handelt es sich bei der Evaluation der Programmwirkungen bei den Adressaten um eine isolierte Evaluation. Hier wird geprüft, ob ein interessierendes Programm überhaupt eine Wirkung hervorbringen kann (vgl. Hager, 1998). Zur Evaluation der Programmwirkungen bei den Adressaten wurde ein quasi-experimentelles Design mit unbehandelter Kontrollgruppe implementiert. Allerdings schien eine Überprüfung der Adressatenwirkung durch einen quasi-experimentellen Versuchsplan nur innerhalb der Schule realisierbar. Hier schienen die Chancen für einen wiederholten Zugang zur Adressatenstichprobe gut. Bei Trainingsvariante 2, die im Setting „Freizeit“ angesiedelt war, wurde davon ausgegangen, dass die Möglichkeiten, Jugendliche zweimal zur Durchführung einer Befragung zu motivieren, sehr viel geringer sein würden. Es war zu erwarten, dass die Datenerhebung schon an dem mangelnden Zugang zur Zielgruppe scheitern würde. Daher beschränkte sich die Evaluation der Programmwirkung bei Adressaten auf die Langzeittrainingsvariante, beim Kurzzeittraining wurde auf eine Evaluation der Programmwirkung bei den Adressaten verzichtet.

	Pretest Vor Beginn der Peeraktivitäten N = 1692	Durchführungsphase: 4-6 Monate	Posttest N = 1411
Berliner Schülerinnen und Schüler	Interventions- Gruppe n = 1094		Interventions- Gruppe n = 924
	Kontroll- Gruppe n = 598		Kontroll- Gruppe n = 487

Abbildung 10-4. Versuchsplan zur Adressatenerhebung

Die Befragung der Adressaten im schulischen Bereich wurde etwa zeitgleich mit der Beendigung der Peer-Educator Ausbildung, noch bevor die Peer-Educators in den Schulen aktiv wurden, durchgeführt. Ein Posttest bei Adressaten unmittelbar im Anschluss an die Peerveranstaltung wäre sehr wünschenswert gewesen, da hier die Chance unmittelbare Treatmentwirkungen festzustellen, am größten war. Dies war jedoch nicht praktikabel, da

die Peer-Educators ihre Veranstaltungen in den Schulen eigenständig organisierten und oft sehr kurzfristig anberaumten. Dies hätte einen hohen organisatorischen Aufwand für die beteiligten Schulen bedeutet, der mit hoher Wahrscheinlichkeit von vorne herein abgelehnt worden wäre. So hätte jeweils im unmittelbaren Anschluss an den Peerunterricht die nachfolgende Unterrichtsstunde ausfallen müssen. Im Falle geplanter Klassenarbeiten wären hier hohe Widerstände von seiten der Lehrer und Schulleitungen zu befürchten gewesen. Zu einer Zusammenfassung der Probleme der Implementation und Evaluation schulischer Interventionsprogramme soll auf Meyer, Miller und Herman (1993) verwiesen werden.

Daher wurde der Posttest aus pragmatischen Gründen in allen Schulen standardisiert nach einem halben Jahr durchgeführt. Dieser Zeitraum wurde festgelegt, um die Möglichkeit eines einigermaßen standardisierten Vergleichs mit den Kontrollschulen zu haben. Die Durchführung dieser Befragung war unabhängig davon, ob es zu Aktivitäten seitens der Peer-Educators gekommen war oder nicht. Dadurch ergab sich die zusätzliche Chance, Diffusionseffekte (vgl. Rogers, 1983) zu testen. Diese sollten sich innerhalb der Interventionsschulen bei Klassen ohne Informationsveranstaltung zeigen. Der Zeitpunkt der Durchführung der Informationsveranstaltung durch Multiplikatoren streute innerhalb der halbjährigen Durchführungsphase sehr stark. In manchen Fällen wurde die Veranstaltung schon zu Beginn der Implementationsphase durchgeführt, in anderen Fällen am Ende. In den meisten Fällen lagen die durchgeführten Peerveranstaltungen häufig schon Monate zurück. Daher erscheint es gerechtfertigt, hier von Langzeiteffekten zu sprechen.

Zusätzlich wurde im unmittelbaren Anschluss an die Informationsveranstaltung noch eine weitere Kurzbefragung mittels eines zweiseitigen Fragebogens durchgeführt. Diese Befragung diente der Feststellung von unmittelbaren Programmeffekten. Diese Kurzbefragung wurde von den Peer-Educators selbst organisiert und dauerte im allgemeinen nur fünf Minuten. Der Kurzfragebogen wurde unmittelbar nach dem Ausfüllen der Evaluationsgruppe postalisch zugesandt.

Trotz der genannten prinzipiellen Schwierigkeiten des experimentellen Forschungsparadigmas im Bereich der angewandten Sozialforschung, erschien uns die Anwendung eines quasi-experimentellen Designs im Augenblick als beste verfügbare Methode, um unter Berücksichtigung der genannten Kritikpunkte zu verallgemeinerbaren Aussagen zu gelangen. Ein Nachweis von Programmwirkungen bei Adressaten erschien uns vor allem auch deswegen besonders bedeutsam, weil dieser Bereich in bisherigen Forschungsarbeiten vernachlässigt wurde. Erst wenn auch Adressaten von einem Peer-Education-Programm profitieren, können Peer-Education-Programme im eigentlichen Sinne als wirksame Multiplikatorenstrategie betrachtet werden. Als eine generelle Schwierigkeit muss dabei anerkannt werden, dass hier im Gegensatz zu dem intensiven Training der Multiplikatoren die Adressaten eine vergleichsweise „schwache“ Intervention erfahren. Im allgemeinen besteht diese in dem einmaligen „Besuch“ einer Peerveranstaltung oder Peeraktivität. Diese Intervention „konkurriert“ beim Nachweis einer längerfristigen Wirkung mit vielfältigen weiteren Informationseinflüssen und Erfahrungen. Falls sich über einen längeren Zeitraum überhaupt Effekte nachweisen

lassen, dürften diese nur sehr gering sein. Unter Effizienzgesichtspunkten wäre es wünschenswert, wenn die potentiell geringe Wirksamkeit durch eine möglichst große Programmreichweite kompensiert werden würde.

10.4.2. DIE DURCHFÜHRUNG DER BEFRAGUNG

Die gesamte Datenerhebung zur Adressatenbefragung wurde innerhalb des Schuljahres 1996/1997 durchgeführt. Dieses Vorgehen empfahl sich zum einen, da es aus Erfahrung schulischer Präventionsprojekte manchmal Schwierigkeiten bereitet, Jugendliche über verschiedene Schuljahre hinweg als Befragungsperson wiederzugewinnen. Zum anderen sind auch Entwicklungseinflüsse, die mit dem Schuljahreswechsel assoziiert sind, nicht zu vernachlässigen, vor allem, da sie meist die Untersuchungsstichprobe nicht gleichermaßen betreffen.

Die Pretesterhebungen begannen im August 1996 und setzten sich bis Oktober 1996 fort. Etwa 6 bis 8 Monate nach der Erstbefragung wurde eine Wiederholungsbefragung durchgeführt. Die ersten Posttests begannen entsprechend Ende Februar 1997 und wurden sukzessiv bis zum Ende des Berliner Schuljahres (Mitte Juni 1997) abgeschlossen.

Ein versuchsplanerisches Problem der Evaluation ergab sich dadurch, dass es nicht möglich war, vor Beginn der Erhebungen die Zielgruppe auf bestimmte Schulklassen einzugrenzen. Der Modellprojektgedanke beinhaltet die freiwillige Zustimmung von seiten der Adressaten, aber auch der Peer-Educators Veranstaltungen und Aktivitäten zum Thema Sexualaufklärung miteinander durchzuführen. Die praktische Umsetzung des Modellprojektes erlaubte es daher leider nicht, die Termine und Schulklassen langfristig festzulegen.

Um nicht das Risiko einzugehen bereits beim ersten Befragungstermin Effekte des Modellprojekts zu erfassen, wie sie zum Beispiel durch Werbeveranstaltungen oder auch durch schon gelaufene Veranstaltungen (Informationsdiffusion) an der Schule erzielt werden können, wurden gesamte Jahrgänge vor Beginn der Durchführungsphase befragt. Die Auswahl der potentiellen Zielgruppe (Adressaten) erfolgte in Absprache mit dem pädagogischen Team (Trainer). So wurde davon ausgegangen, dass die Peer-Educators eine Altersgruppe bevorzugen würden, die ein, maximal zwei Jahrgänge unter ihrer eigenen Altersgruppe liegen würde. Die Befragungsmöglichkeiten der Adressaten waren vor allem auch dadurch beschränkt, dass eine Genehmigung zur Befragung der Schülerinnen und Schüler von der Senatsverwaltung für Schule, Jugend und Sport nur für die achten bis zehnten Klassen erteilt wurde. In Abhängigkeit von dem Durchschnittsalter der Peer-Educator - Gruppen und ihrer potentiellen Zielgruppe wurden alle achten bis zehnten Klassen aus zwei Gymnasien, der achte und neunte Jahrgang aus einer Gesamtschule und vier achte Klassen einer Hauptschule befragt.

Abbildung 10-4 verdeutlicht, dass die Zahl der Interventionsschüler im Vergleich zu den Kontrollschülern deutlich höher ist. In der Hauptsache ist die geringere Schülerzahl der Kontrollschulen dafür verantwortlich. Die Auswahl der Schulen für das Modellprogramm erfolgt nicht mittels Zufallsauswahl, sondern war durch schon

vorhandene soziale Kontakte des Praxisteam zu Schulen und die Bereitschaft der Schulen zur Zusammenarbeit bedingt. Positive Selektionsprozesse bei der Auswahl der Interventionsschulen sind nicht auszuschließen.

Die längsschnittliche Stichprobe, die den folgenden statistischen Analysen zugrunde liegt, umfasst insgesamt 1411 Jugendliche, davon sind 55.2 Prozent Mädchen und 44.8 Prozent Jungen. Die Zahlen repräsentieren diejenigen Jugendlichen, die zweimal - also im Längsschnitt - befragt werden konnten. In Wirklichkeit wurden jedoch bei der ersten Befragung nicht 1411 Schülerinnen und Schüler befragt, sondern 1692. Davon konnten 83.4 Prozent zum zweiten Messzeitpunkt (N=1411) wiedergewonnen werden. Der Verlust an Befragungsteilnehmern ist in den Kontrollschulen etwas höher mit 18.6 Prozent versus 15.5 Prozent in den Interventionsschulen.

Das Durchschnittsalter beim ersten Messzeitpunkt liegt bei 14.3 Jahren, die Jüngsten sind zum Zeitpunkt der Erstbefragung 12 Jahre, die Ältesten 17 Jahre alt. Eine weitere Beschreibung der Adressaten hinsichtlich ihrer sexuellen Erfahrung und ihres Verhütungsverhaltens erfolgt im Ergebnisteil.

10.4.3. MESSINSTRUMENTE

Das wesentlichste Ziel der Entwicklung eines Fragebogens zur Ergebnisevaluation ist die Operationalisierung von Interventionszielen und Erfolgskriterien. Ein wesentliches Ziel für die Trainer war die Förderung der Kommunikation zu Sexualität und Verhütung, das zweite übergreifende Ziel die Förderung des Kondomgebrauchs bei Jugendlichen. Es werden im Folgenden nur diejenigen Messinstrumente und Indikatoren vorgestellt, die für die dargestellten Analysen bedeutsam sind. Bei der Konzeption des Fragebogens wurde Wert darauf gelegt, die Fragen so zu formulieren, dass sie auch bei einer Gruppe Hauptschüler, die in die Untersuchung mit einbezogen wurde, zu keinerlei Verständnisschwierigkeiten führen würde. Bezüglich der Zahl zu erfassender Variablen wurde zugunsten der Messgenauigkeit (Fidelitäts-/ Bandbreitendilemma, vgl. Cronbach & Gleser, 1965) und zuungunsten einer Vielzahl potentiell interessierender Variablen entschieden. Nach unserer Erfahrung erzeugen zu lange Fragebogen Reaktanz und Ermüdungserscheinungen bei Jugendlichen, so dass nur zentral erscheinende Indikatoren erfasst wurden. Der Fragebogen sollte problemlos von allen beteiligten Schülern innerhalb einer Schulstunde zu beantworten sein.

I Der erste Variablenkomplex umfasst Items mit hoher Inhaltsvalidität, die sich auf den Bereich der Kommunikationsförderung beziehen. Dieser Bereich wurde miteinbezogen, da ein erklärtes Ziel der Trainer eine Steigerung der Kommunikation und Auseinandersetzung der Jugendlichen mit Themen wie Liebe, Sexualität und Schwangerschaftsverhütung war. Für dieses spezifische Modellprojektziel mussten neue möglichst verhaltensnahe Items, bzw. Skalen entwickelt werden. So wurde die selbsteingeschätzte Kommunikationshäufigkeit mit der Frage erfasst: „Wie häufig hast Du in den letzten sechs Monaten über die folgenden Themen gesprochen?“. Das gewählte Zeitfenster von sechs Monaten war an die Dauer der Durchführungsphase adaptiert

(Antwortmöglichkeiten 1 „fast jeden Tag“ 2 „mehrmals pro Woche“ 3 „mehrmals im Monat“ 4 „mehrmals im halben Jahr“ 5 „seltener bzw. nie“). Durch Verkürzung von ursprünglich zwölf auf drei Items wurde ebenso wie bei den Peer-Educators die Kommunikationsschwierigkeit erfragt (Beispielitem: „Wie schwierig wäre es im Moment für Dich, mit anderen Jugendlichen offen über Sexualität zu reden“). Durch drei identische Items wurde (Beispielitem: „Ich finde es wichtig, mit anderen Jugendlichen offen über Sexualität zu reden“) die Kommunikationswichtigkeit erfragt. Diese kann als ein Indikator der Motivation betrachtet werden. Die internen Konsistenzen der drei Skalen zum ersten und zweiten Messzeitpunkt sind zufriedenstellend, die Retest - Reliabilität erscheint ebenfalls als ausreichend (vgl. Tabelle 10-5), wenn man berücksichtigt, dass das Jugendalter als eine Phase gravierender entwicklungsbedingter Veränderungen zu betrachten ist.

Als weitere Indikatoren der Motivation wurden die Einstellung und die Intention zur Kommunikation anhand von single indicators erfasst. Das Item zur Erfassung der Einstellung lautet: „Mit meinen Freunden über Sexualität zu sprechen, finde ich ...“, der Abfragemodus war siebenstufig von -3 „sehr schlecht“ über 0 bis +3 „sehr gut“. Analog dazu wurde die Intention anhand der Frage erfasst „Ich habe die Absicht, ab morgen mit meinen Freunden/Freundinnen offen über Sexualität zu sprechen“. Die Antwortmöglichkeiten waren ebenfalls siebenstufig von -3 „ganz unwahrscheinlich“ bis +3 „sehr wahrscheinlich“. Nach Schwarzer (1996, S. 3) soll sich die in Frage stehende Intention zum „kommunizieren“ auf eine unmittelbar bevorstehende Handlung beziehen. Dies begründet sich darauf, dass die Stabilität einer Intention geringer wird, wenn der Zeitabstand zum Verhalten groß ist. Verschiedene Barrieren stehen der Umsetzung des Verhaltens entgegen. Dabei besteht die Gefahr, dass sich der angenommene Zusammenhang zwischen Intention und Verhalten immer mehr reduziert, je länger der Zeitabstand zwischen Intention und Verhalten ist. Wie die Tabelle 10-5 zeigt, sind die Stabilitäten der kommunikationsbezogenen Einstellung und Intention sehr gering, nur etwa 16 bis 18 Prozent an Varianz zwischen erstem und zweitem Messzeitpunkt überlappen. Hierfür mag die mangelnde Reliabilität der Items verantwortlich sein.

Tabelle 10-5. Mittelwerte, Interne Konsistenz und Retest – Reliabilität der Instrumente des Adressatenfragebogens zum Bereich Kommunikation

	Mittelwert (SD) der Skala	Interne Konsistenz Cronbach's Alpha	Retest-Reliabilität r _{tt}
Bereich Kommunikation			
Einstellung zur Kommunikation (Einzelindikator, Range 1-7)	aM = 5.68 (1.32)		r _{tt} = .43
Intention zur Kommunikation (Einzelindikator, Range 1-7)	aM = 4.22 (1.97)		r _{tt} = .45
Wahrgenommene Kommunikationsschwierigkeit (3-Itemskala, Range 1-4)	aM = 2.66 (0.75)	$\alpha = .79$	r _{tt} = .58
Wahrgenommene Kommunikationswichtigkeit (3-Itemskala, Range 1-4)	aM = 1.93 (0.71)	$\alpha = .79$	r _{tt} = .49
Selbstwirksamkeitserwartung, Initiative ergreifen zu können (im sexuellen Bereich) (3-Itemskala, Range 1-5)	aM = 3.27 (0.89)	$\alpha = .65$	r _{tt} = .53
Kommunikationshäufigkeit (4-Itemskala, Range 1-5)	aM = 2.91 (0.85)	$\alpha = .74$	r _{tt} = .61

II Als zweites Erfolgskriterium wurde die Förderung des Kondomgebrauchs betrachtet. Allerdings wurde der Kondomgebrauch zum ersten Messzeitpunkt nicht direkt erfragt, da dies bei der Senatsverwaltung für Schule, Jugend und Sport auf Widerstand stieß. Erst beim zweiten Messzeitpunkt war dies möglich, ebenso wie das Erfragen sexueller Erfahrungen. Es wurden jedoch Kognitionen und Einstellungen erfragt, die nachgewiesenermaßen mit dem Kondomgebrauch in Zusammenhang stehen sollten. Als theoretische Basis diente die Theory of Planned Behavior (Ajzen & Madden, 1986). Zur Operationalisierung der entsprechenden Konstrukte wurde ad hoc eine Reihe von Items und Kurzskalen entwickelt.

Zum einen wurde identisch mit der Erfassung von Einstellung und Intention zur Kommunikation auch die Einstellung und Intention zum Kondomgebrauch operationalisiert. Die Korrelation zwischen den beiden Messzeitpunkten sind ähnlich gering wie bei der Kommunikation. Darüber hinaus muss die Einstellung der Jugendlichen zum Kondomgebrauch als ausnehmend positiv charakterisiert werden bei einem Mittelwert von 6.09 (Range: 1 bis 7), ebenso ist die Intention zum Kondomgebrauch zum ersten Messzeitpunkt stark ausgeprägt mit einem Mittelwert von 5.57 (Range: 1 bis 7). Die Verteilung der Items ist entsprechend rechtsschief (vgl. Anhang). Aufgrund des Deckeneffektes sollte das Potential, sich vom ersten zum zweiten Messzeitpunkt in positive Richtung zu verändern, beschränkt sein. Die Einstellung ist das Resultat einer Mehrzahl subjektiver Überzeugungen zu Konsequenzen des Kondomgebrauchs. Es wurden

drei positive und vier negative Konsequenzerwartungen zum Kondomgebrauch erfragt. Ein Beispielitem für eine positive Konsequenzerwartung lautet: „Wenn ich Kondome benutze, dann schütze ich mich vor Aids“, ein Beispiel für eine negative Konsequenzerwartung „Wenn ich Kondome benutze, dann stört das die Spontaneität beim Sex“. Der Abfragemodus war fünfstufig. Die sieben Items wurden zwecks Skalenbildung einer Faktorenanalyse unterzogen und luden, wie zu erwarten auf je einem Faktor positiver und negativer Konsequenzerwartungen. Die internen Konsistenzen waren für die negativen Konsequenzerwartungen recht niedrig und für die positiven Konsequenzerwartungen so gering, dass diese von den weiteren Analysen vollständig ausgeschlossen werden mussten. (vgl. Tabelle 10-6). Neben Einstellung, Konsequenzerwartungen und Intention stellt auch die sogenannte subjektive Norm ein wesentliches Bestimmungsstück der Theory of Planned Behavior dar. Die subjektive Norm widerspiegelt das, was Personen, die einem wichtig sind („the significant other“), erwarten. Diese Erwartung sollte möglichst spezifisch formuliert sein und im Spezifitätsgrad der vorherzusagenden Intention entsprechen. Normative Überzeugungen bringen den generalisierten Erwartungsdruck anhand konkreter Bezugspersonen zum Ausdruck: Normative Überzeugungen zum Kondomgebrauch wurden in Bezug auf Eltern, Freunde und Lehrer erfragt (Beispielitem: „Meine Eltern erwarten von mir sicherlich, dass ich Kondome benutzen würde“) und bildeten eine Gesamtskala. Auf die Erfassung der „motivation to comply“, das heißt der Bereitschaft den Erwartungen zuzustimmen, wurde verzichtet.

In den „neueren“ gesundheitspsychologischen Modellen wie zum Beispiel dem Limburger Modell (DeVries, Dijkstra & Kuhlmann, 1988) wird auch die Selbstwirksamkeitserwartung in die Verhaltensvorhersage mit einbezogen. In der vorliegenden Untersuchung wurden verschiedene Aspekte der Selbstwirksamkeitserwartung berücksichtigt, beispielsweise als Zutrauen in die Fähigkeit, im sexuellen Bereich Initiative ergreifen zu können, oder aber auch den Schutz mit Kondomen in einer intimen Situation durchzusetzen.

Die neun Items (Beispielitem: „Ich traue mir zu, das Kondom/Verhütungsmittel richtig zu benutzen“) wurden mittels einer Faktorenanalyse (Hauptkomponentenanalyse mit schiefwinkliger Rotation) auf zwei Faktoren reduziert. Um das Kriterium der „relativen Eindimensionalität“ nicht zu verletzen, wurden zwei Items des ursprünglichen Itempools ausgeschlossen. Dieses Kriterium besagt, dass mehrfach ladende Items eine Differenz von mindestens .30 in der Ladungshöhe haben sollten. Der erste Faktor („Initiative ergreifen“) bezieht sich auf die Fähigkeit, eigene Bedürfnisse und Wünsche auszudrücken und wurde dem Bereich der Kommunikation zugeordnet (vgl. Tabelle 10-5). Beim zweiten Faktor hingegen („Kontrolle ausüben“) geht es eher um die Fähigkeit, Sexualverhalten zu hemmen und kontrollieren. Diese Fähigkeit ist auch auf die Interaktion mit dem Sexualpartner bezogen (vgl. Tabelle 10-6) und schließt die Fähigkeit zum „Nein-sagen“ ein. Reliabilitäten sind in beiden Fällen im unteren Bereich des noch Akzeptablen. Als letztes Item wurde die wahrgenommene Barriere des Kondomgebrauchs erfragt (Frage: „Was glaubst Du, wie schwer es Dir fallen würde, ein Kondom zu benutzen?“). Die überwiegend niedrigen Stabilitäten dürften durch die geringen internen Konsistenzen mitbedingt sein. Nicht auszuschließen ist hierbei, dass Messprobleme dabei eine Rolle

spielen. Andererseits sind jedoch Veränderungen im Sinne einer Merkmalsveränderung auch intendiert und gerade im Jugendalter mehr noch als in anderen Entwicklungsphasen zu erwarten.

Tabelle 10-6. Mittelwerte, Interne Konsistenzen und Retest-Reliabilität der Instrumente des Adressatenfragebogens zum Bereich Kondomgebrauch

	Mittelwert (SD) der Skala	Interne Konsistenz Cronbach's Alpha	Retest-Reliabilität r_{tt}
Bereich Kondomgebrauch			
Einstellung zum Kondomgebrauch (Einzelindikator; Range 1-7)	aM = 6.09 (1.29)		$r_{tt} = .48$
Positive (Konsequenz)-Erwartungen zum Kondomgebrauch (3-Itemskala, Range 1-5)	aM = 4.67 (0.53)	$\alpha = .55$	$r_{tt} = .34$
Negative (Konsequenz)-Erwartungen zum Kondomgebrauch (4-Itemskala, Range 1-5)	aM = 1.74 (0.74)	$\alpha = .68$	$r_{tt} = .48$
Normative Erwartungen der Eltern zum Kondomgebrauch (Einzelindikator, Range 1-6)	aM = 5.06 (1.35)		$r_{tt} = .50$
Normative Erwartungen der Lehrer zum Kondomgebrauch (Einzelindikator, Range 1-6)	aM = 4.28 (1.66)		$r_{tt} = .39$
Normative Erwartungen der Freunde zum Kondomgebrauch (Einzelindikator, Range 1-6)	aM = 4.99 (1.39)		$r_{tt} = .37$
Normative Erwartung Gesamtskala (3-Itemskala, Range 1-6)	aM = 4.78 (1.16)	$\alpha = .72$	$r_{tt} = .48$
Selbstwirksamkeitserwartung zum Ausüben von Kontrolle (4-Itemskala, Range 1-5)	aM = 4.24 (0.77)	$\alpha = .70$	$r_{tt} = .47$
Barriere der Kondombenutzung (Einzelindikator, Range 1-7)	aM = 2.49 (1.52)		$r_{tt} = .45$
Intention zum Kondomgebrauch (Einzelindikator, Range 1-7)	aM = 5.57 (1.98)		$r_{tt} = .38$

III Darüber hinaus wurden noch drei Subskalen des Berner Fragebogens zum Wohlbefinden (Grob et al., 1989) in den Adressatenbogen mit aufgenommen: Die positive Lebenseinstellung, das Selbstwertgefühl und Körperliche Beschwerden. Diese können im Rahmen stresstheoretischer Modellvorstellungen auch als personale und soziale Bewältigungsressourcen betrachtet werden.

Tabelle 10-7. Mittelwerte, Interne Konsistenzen und Retest-Reliabilität der Instrumente des Adressatenfragebogens zum Bereich Wohlbefinden

	Mittelwert (SD) der Skala	Interne Konsistenz Cronbach's Alpha	Retest-Reliabilität r _{tt}
Wohlbefinden			
BFW-Subskala: Positive Lebenseinstellung (8-Itemskala; Range 1-6)	aM = 4.3 (.73)	$\alpha = .80$	r _{tt} = .61
BFW-Subskala: Selbstwert (3-Itemskala; Range 1-6)	aM = 4.7 (.85)	$\alpha = .74$	r _{tt} = .57
BFW-Körperliche Beschwerden (9-Itemskala; Range 1-6)	aM = 1.93 (.57)	$\alpha = .72$	r _{tt} = .64

Zu ergänzen bleibt, dass als ein Erfolgskriterium mit hoher Inhaltsvalidität bei Adressaten die Kenntnis von Beratungsstellen erfragt wurde. Gefragt wurde: „Kennst Du eine oder mehrere Beratungsstellen für die Probleme von Jugendlichen?“ Die Antwortalternativen lauteten: 1 „keine“ 2 „ja, eine“ 3 „ja, mehrere“. Im Anschluss wurden die Jugendlichen gebeten, alle Beratungsstellen, die sie kannten, aufzuschreiben. Die Anforderung der aktiven Erinnerung von Beratungsstellen erhöht die Validität dieses Items.

Schließlich wurde noch ein Item zur Überprüfung des Kenntnisstandes des Konzeptionszeitpunktes, d.h. des Zeitpunktes der höchsten Fruchtbarkeit bei Mädchen („Konzeptionsoptimum“) eingefügt. Eine Vermittlung von Wissen war in dem Programm nicht geplant und wäre somit eine nichtbeabsichtigte Programmwirkung. Andererseits dürfte die Kenntnis des Zeitpunktes der höchsten Fruchtbarkeit bei Mädchen („Konzeptionsoptimum“) eine wichtige Rolle für den Verhütungsmittelgebrauch spielen. Daher wurde ein solches Item aus dem Fragebogen von Eichholz et al. (1994) hier eingesetzt. Das Item lautet „Was meinst Du, wann ist die Möglichkeit, schwanger zu werden, am größten?“ Mögliche Antwortalternativen waren: „etwa in der Mitte zwischen zwei Regelblutungen“, „in der ersten Tagen während der Regelblutung“, „in den ersten Tagen nach der Regelblutung“ und „in den letzten Tagen vor der nächsten Regelblutung“.

Zur Erfassung von Kurzzeiteffekten wurde unmittelbar nach jeder Veranstaltung ein zweiseitiger Kurzfragebogen eingesetzt, der dazu dient, unmittelbare Effekte der Intervention zu erfassen. Der Feedbackbogen beinhaltet unter anderem eine Bewertung der Veranstaltung und der Präsentationsfähigkeiten der Peer-Educators, sowie 6 ausgewählte Zielkriterien des Programms, die auch im Pretest erhoben wurden. Aus dem Pretestfragebogen wurden Einstellungen und Intentionen zur Kondombenutzung und zur Kommunikation als single indicators übernommen. Ebenso wurden die wahrgenommene Kommunikationsschwierigkeit (Cronbach's $\alpha = .78$) und Kommunikationswichtigkeit (Cronbach's $\alpha = .85$) übernommen.

10.4.4. BESCHREIBUNG DER INTERVENTION/TREATMENT

Entgegen der anfänglichen Planung, Peeraktionen nur im Freizeitbereich anzubieten, wurden auch im schulischen Bereich Peerveranstaltungen durchgeführt, die sich dem Typ Peeraktion zuordnen lassen. So bestand die Möglichkeit zwei Veranstaltungstypen in ihrer Wirkung zu evaluieren. Erstens eine Informationsveranstaltung der Peer-Educators in Schulklassen und zweitens ein Theaterstück, das vor einem größeren Publikum - meist vor Schulklassen -aufgeführt wurde.

10.4.4.1. Aktivitäten der Peer-Educators in Schulklassen

Die Informationsveranstaltungen der Jugendlichen an Schulen wurden meist innerhalb einer oder zwei Schulstunden durchgeführt. Im allgemeinen führten die Multiplikatoren diese Sitzung in der Schulklasse zu zweit durch. Aus formalen Gründen (Schulaufsicht) war es erforderlich, dass die Peer-Educators an ihrer eigenen Schule aktiv wurden. Die Aktivitäten waren vielfältig und reichten von der Demonstration der Benutzung von Verhütungsmitteln und Kondomen bis hin zu Spielen, die zur Reflektion von Werten und Einstellungen anregen sollten. Das Programmkonzept, das für Peer-Educators die freie Auswahl von Methoden in einer Sitzung vorsah, erlaubte nicht die Evaluation einzelner Interventionselemente, sondern lies nur Aussagen über die summative Wirkung der gesamten Veranstaltung zu. Als kleinster gemeinsamer Nenner wurde in allen Interventionsklassen über die hier berichtet wird, eine Verhütungsmittel- und Kondomdemonstration als mindestens ein gemeinsames Element der Sitzung durchgeführt. In Tabelle 10-8 ist die Anzahl der insgesamt durchgeführten und in der Langzeitevaluation berücksichtigten Informationsveranstaltungen im betrachteten Durchführungszeitraum dargestellt.

Tabelle 10-8. Anzahl der durchgeführten und in der Langzeitevaluation berücksichtigten Informationsveranstaltungen in einem Zeitraum von 6-8 Monaten

	Zahl durchgeführter Informations-Veranstaltungen	Zahl ausgewerteter Informations-veranstaltungen	Begründung für die Nichtberücksichtigung
Gymnasium 1	7	7	
Gymnasium 2	7	7	
Gesamtschule	8 ^a	4 ^b	4 Informationsveranstaltungen bei Siebtklässlern ^c
Hauptschule	10	4	6 Informationsveranstaltungen bei Siebtklässlern ^c

Anmerkung. ^a in zwei Schulklassen wurden zwei Informationsveranstaltungen durchgeführt; ^b in einer Schulklasse wurden zwei Informationsveranstaltungen durchgeführt ; ^c Siebtklässler konnten aufgrund der fehlenden Bewilligung der Befragung der Schülerinnen und Schüler durch die Senatsverwaltung für Schule, Jugend und Sport nicht in die Evaluation mit einbezogen werden

10.4.4.2. Theaterstück der Peer-Educators zum Thema „Bisexualität“

Bei einer gymnasialen Multiplikatorengruppe, die in den beiden Gymnasien (Gymnasium 1 und Gymnasium 2) aktiv wurde, wurde zusätzlich zur Vorbereitung auf die Informationsveranstaltungen noch ein Theaterstück zum Thema „Bisexualität“ einstudiert. Hier ging es um die Auseinandersetzung mit gesellschaftlichen Normen und Vorurteilen gegenüber bi- und homosexuellen Menschen. Das Stück war im zweiten Teil als eine Art Talkshow angelegt, in die das Publikum möglichst aktiv miteinbezogen wurde. Es wurde an beiden Schulen in der Aula vor den achten bis zehnten (bzw. achten bis elften) Klassen aufgeführt. In der anschließenden Diskussion kam es teilweise zu heftigen Kontroversen und Meinungsverschiedenheiten. An beiden Gymnasien wurde von der Schulleitung ein Vormittag für das Theaterstück zur Verfügung gestellt. Das Stück selbst dauerte mit anschließender Diskussion etwa 1 bis 1 ½ Stunden und wurde an dem betreffenden Vormittag vor wechselndem Publikum drei- bis vier Mal wiederholt.

Bei beiden Gymnasien wurde das Theaterstück noch vor den Informationsveranstaltungen durchgeführt und erreichte im wesentlichen alle Adressaten, die in die Evaluation mit eingeschlossen worden waren, so dass am Ende nur Aussagen über den summativen Effekt beider Treatments - Theaterstück und Informationsveranstaltung – zugelassen sind.

10.5. Statistische Auswertung

Die Auswertung des quantitativen Untersuchungsteils wurde mittels SPSS 9.0 für Windows vorgenommen. Zunächst erfolgte eine Datenbegutachtung. Die gebildeten Skalen wurden hinsichtlich der Verteilung pro Gruppe und Messzeitpunkt per Augenschein begutachtet. Die Überprüfung der Normalverteilung erfolgte anhand der Schiefe und der

Durchführung von Kolmogorov – Smirnov Tests pro Skala, bzw. single indicator und Messzeitpunkt. Information zur Schiefe ist wie bereits erwähnt im Anhang angegeben. Auf die Verletzung der Normalverteilungsannahme wird im Zusammenhang mit der Ergebnisauswertung aufmerksam gemacht.

Zur Überprüfung von Programmwirkungen bei Multiplikatoren wurden jeweils Varianzanalysen mit Messwiederholung (Prozedur: Manova) gerechnet. Bei der Feststellung von Programmwirkungen bei Multiplikatoren mit Langzeittraining wurde eine Messwiederholungsanalyse über drei Messzeitpunkte gerechnet, zunächst ohne Gruppierungsfaktoren zu berücksichtigen. Um Aussagen nicht nur bezüglich der globalen Veränderungseffekte (T1 bis T3) treffen zu können, wurden im zweiten Schritt die signifikanten Ergebnisse anhand spezieller Kontraste überprüft. Der Kontrast „Repeated“ vergleicht jeweils die erste Stufe eines Faktors mit der darauffolgenden Stufe. Dies ermöglicht den Vergleich von T1 mit T2 und ebenso von T2 mit T3. Allerdings sind die so berechneten Kontraste nun nicht mehr orthogonal, da T2 in beide Analysen mit eingeht.

Beim Vergleich der Trainingsvarianten ging jeweils ein Gruppierungsfaktor mit zwei (1 Berliner Langzeittraining 2 Berliner Kurzzeittraining) bzw. drei (1 Berliner Langzeittraining 2 Grazer Kurzzeittraing 3 Grazer Kontrollgruppe) Stufen als unabhängiger Faktor in die Auswertung mit ein. Als abhängige Variable diente der dreistufige oder zweistufige Messwiederholungsfaktor.

Die Auswertung der Programmwirkungen bei Adressaten erfolgte ebenfalls durch Varianzanalyse mit Messwiederholung. Als abhängige Variable diente der Messwiederholungsfaktor mit zwei Stufen und als unabhängige Variable ein zweistufiger Gruppierungsfaktor – Interventions- versus Kontrollschule. Die unterschiedlichen Aktivitäten der Peer-Educators in verschiedenen Schulen erschwerte die Interpretation einer Globalauswertung der Programmwirkungen in einer einzigen statistischen Analyse. Daher wurden im Anschluss an die Globalauswertung Schulen mit vergleichbarem Treatment zusammengefasst und in Subgruppen analysiert. Schließlich bot sich für eine Subgruppe von Gymnasiasten die Gelegenheit, auch Kurzzeitwirkungen in die längsschnittliche Analyse mit einzubeziehen, so dass auch hier Messwiederholungsvarianzanalysen mit 3 Messzeitpunkten berechnet wurden. Als Gruppierungsfaktor ging ein Treatmentfaktor mit 2 Stufen (1 schwache Intervention 2 starke Intervention) in die Auswertung mit ein.

Zum Vergleich der Kontrollgruppe mit Interventionsgruppe hinsichtlich potentiell wirksamer Störvariablen, der Trainingsaussteiger mit Dabeibleibern und der Trainingsbewertung durch Multiplikatoren mit unterschiedlichem Treatment wurden ein-, bzw. zweifaktorielle Varianzanalysen durchgeführt.

Um Ergebnisse statistischer Verfahren wie Varianzanalysen statistisch interpretieren zu dürfen, müssen bestimmte Voraussetzungen der Daten erfüllt sein, wie sie beispielsweise von Bortz (1993, S. 261) beschrieben wurden. Als erste Voraussetzung gilt, dass die Fehlerkomponenten in den Grundgesamtheiten, denen die Stichprobe entnommen wurde normalverteilt sein müssen („normalverteilte Fehlerkomponenten“). Diese Annahme wird als erfüllt angesehen, wenn die Variable pro Treatmentstufe normalverteilt ist. Die zweite Annahme - homogene Fehlervarianzen - beinhaltet, dass die Stichproben aus

Grundgesamtheiten stammen, in denen die Messwerte die gleiche Varianz aufweisen. Die Varianzen innerhalb der Stichproben dürfen sich daher nicht signifikant unterscheiden. Diese Voraussetzung wird unter anderem mit dem Bartlett Test überprüft. Allerdings führt der Bartlett-Test nur dann zu richtigen Entscheidungen, wenn die Normalverteilungsannahme erfüllt ist. Als ein robusteres Verfahren wird der Levené- Test beschrieben, der in den folgenden Analysen auch eingesetzt wurde. Die dritte Annahme - die „unabhängigen Fehlerkomponenten“ meint, dass keine systematische Beeinflussung eines Messwertes durch Störvariablen vorhanden sein soll. Diese wird als gegeben angesehen, wenn eine randomisierte Zuordnung der Untersuchungseinheiten zu Treatmentstufen erfolgt. Im Falle eines quasi-experimentellen Designs ist diese allerdings nicht gegeben.

Wie später noch deutlich werden wird, sind die Voraussetzungen der Varianzanalyse nicht in jedem Fall erfüllt. Jedoch verlieren die Voraussetzungen der Varianzanalyse mit wachsendem Umfang der untersuchten Stichproben an Bedeutung. Daher sollten die Verletzung von Voraussetzungen der Varianzanalyse im Falle der Adressatenerhebung von geringerer Bedeutung als in der Multiplikatorenstichprobe sein.

Für die Varianzanalyse mit Messwiederholung gelten spezielle Voraussetzungen. Hier ist die Annahme der Unabhängigkeit der Fehlerkomponenten ja zwangsweise verletzt, da dieselben untersuchten Personen unter mehreren Treatmentstufen beobachtet werden. In Varianzanalysen mit Messwiederholung müssen die Varianzen unter den einzelnen Faktorstufen und die Korrelationen zwischen den Faktorstufen homogen sein. Diese Voraussetzung wird häufig auch als Homogenität der Varianz-Kovarianzmatrix bezeichnet, und wird per Voreinstellung der Prozedur Manova mit dem Mauchly Sphericity Test überprüft. (Fehres & Marshall, 1990). Da diese Voraussetzung in Messwiederholungsdesigns in der Regel verletzt ist, erfordert die Varianzanalyse fast immer hierauf abgestellte Korrekturverfahren, die die Reduktion der Freiheitsgrade durch den Faktor beinhalten. In der vorliegenden Arbeit wurde das Korrekturverfahren nach Huynfeldt eingesetzt. Dieses basiert auf der Annahme, dass die Differenzen zwischen den jeweils zu zwei Messzeitpunkten erhobenen Varianzen homogen sind.

In der vierten und letzten Fragestellung schließlich ging es um die Bedeutung des Modellernens für zu beobachtende Programmwirkungen. Die Wahrnehmung des Modells als ähnlich, attraktiv oder kompetent sollte einen Einfluss auf gewünschte Programmwirkungen ausüben. Die bereits erwähnte gymnasiale Subgruppe, die über mehrere Messzeitpunkte verfolgt werden konnte, bot die Gelegenheit, den vermittelnden Einfluss von Modellwirkungen zu überprüfen, ohne sich dem Risiko einer zu großen Heterogenität des Treatments auszusetzen. Diese Fragestellung (Fragestellung 4) wurde anhand eines Pfadmodells (Programm: Amos 4.01) überprüft. Hier ging es insbesondere um die angenommene Mediatorwirkung der Modellwahrnehmung für die Programmwirkung.

Schließlich wurde auch die Theory of Planned Behavior einer Überprüfung unterzogen. Auch hierfür wurde eine hypothesenprüfende Pfadanalyse eingesetzt. Die Besonderheit der hypothesenprüfenden im Vergleich zu den explorierenden Verfahren

besteht darin, dass eine gesicherte inhaltliche Theorie über die Zusammenhänge zwischen den Variablen vorliegt. Nach Backhaus (1980, S. 309) sollen zur Durchführung einer Pfadanalyse noch folgende statistische Voraussetzungen gelten:

Die beobachteten Variablen folgen einer Multinormalverteilung. Für das Strukturmodell wird angenommen, dass die Residuen nicht mit den exogenen latenten Variablen korrelieren und die Erwartungswerte der Residuen Null sind. Zusätzlich wird angenommen, dass die Messfehler nicht mit den Residuen der Strukturgleichungen und nicht mit den anderen Konstrukten korreliert sind. Es wird Linearität und Additivität der Konstrukte und Messhypothesen unterstellt. Um die Parameter schätzen zu können, muss die modelltheoretische Kovarianzmatrix positiv definit, d.h. invertierbar sein und das Modell muss identifizierbar sein.